PHYSICAL ORGANIC PRINCIPLES GOVERNING THE SPONTANEOUS PREBIOTIC EMERGENCE OF PROTO-NUCLEIC ACIDS

A Dissertation
Presented to
The Academic Faculty

by

David M. Fialho

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Chemistry and Biochemistry

Georgia Institute of Technology
August 2019

PHYSICAL ORGANIC PRINCIPLES GOVERNING

THE SPONTANEOUS PREBIOTIC EMERGENCE OF

PROTO-NUCLEIC ACIDS

Approved by:

Dr. Nicholas V. Hud, Advisor
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Gary B. Schuster
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Charles L. Liotta
School of Chemistry and Biochemistry
School of Chemical and Biomolecular Engineering
Georgia Institute of Technology

Dr. Stefan France
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Martha Grover
School of Chemical and Biomolecular Engineering
Georgia Institute of Technology

Date Approved: July 22, 2019

ACKNOWLEDGEMENTS

I wish to express profound gratitude to my advisor, Prof. Nicholas Hud. Nick, thank you for taking a chance and hiring me as a technician. It was the best thing for my career that could have happened to me. Nick truly stands out as a research advisor, not only for his intellect and perspicacity, but also in an emphatically pedagogical sense: Nick knows the exact amount of attention each student needs in order to properly flourish, even if that amount changes over time. Early on, he correctly identified my passion in physical organic chemistry, and was enthusiastic about me specializing in it for my Ph.D. He has given me freedom when I wanted it and support when I needed it. Nick, I hope that the tides of our field change soon, and if they do, it will be because of your contributions. I will never forget that.

To Prof. Stefan France, thank you for being my mentor in synthetic organic chemistry. When working in a biophysical chemistry lab, one might not expect to require expertise in organic synthesis, but the knowledge and skills that I gained from Prof. France turned out to be immensely valuable to my research.

To Prof. Martha Grover, thank you for teaching me the value of kinetics in understanding chemical phenomena in a way that elevates organic mechanisms from conjecture to theory.

To Prof. Gary Schuster, thank you for your many lessons on how to be an effective and honest researcher: fail fast, go for the jugular, and remember that we never really know where we're going to need a fire hydrant.

To Prof. Charles Liotta, thank you for being my mentor in physical organic chemistry. Teaching your physical organic chemistry class was one of the most rewarding experiences I had in graduate school. Thank you for the many books you gave me on the fundamentals of physical organic chemistry.

I also wish to thank Prof. Ramanarayanan Krishnamurthy, who has taught me many valuable lessons in prebiotic chemistry and in science communication.

The academic mentors that I had before graduate school inspired me and brought me up from difficult times. Without them, I don't know if I would have the self-esteem to be a scientist. Thank you to Prof. Katherine Masters, my first mentor in organic chemistry, Prof. Jack Vanden Heuvel, my undergraduate research advisor, and Prof. William Noid, for giving me the opportunity to become an undergraduate teaching assistant in thermodynamics, and for teaching me the value of scientific rigor. Thank you to Dr. Nancy Oliveira, who showed me my first prebiotic chemistry experiment, and who inspired me to pursue an academic career studying the origin of life. I hope she reads this one day.

I was incredibly lucky to have a friendly and supportive lab family over the years at Georgia Tech. Thank you so much to Brian Cafferty, Isaac Gállego, Irena Mamajanov, Chelsea Walker, Jaheda Khanam, Denise Okafor, Caterina Musetti, Christine He, Suneesh Karunakaran, Gary Newnam, Bradley Burcar, Jarrod Pellissier, Kimberly Clarke, César Menor-Salván, Adriana Lozoya-Colinas, Martin C, Moran Frenkel-Pinter, Tyler Roche, Chiamaka Obianyor, Kaitlin Jacobson, Sreejith Mangalath, Bryce Clifton, Aikomari Guzmán-Martínez, and Christine Conwell. I would especially like to thank and recognize Isaac Martínez-Aguilar, Megan Moore, and Katherine Watkins, who were all uncommonly

talented undergraduate researchers who I had the pleasure of working with, and who were infinitely patient with me as I learned how to be a mentor.

Without the love, encouragement, and acceptance of my friends, I could not possibly have accomplished this. To Shane Cimino, Evan Andrews, Hunter Norum, Amber Hughes, Adam Rogers, Nina Rao, Kong Wong, Evelyn Ligon, David Hanna, Zachary Duca, Robert Hincape, Rio Febrian, Giorgio Morbioli, Kerry Strickland, and Christopher Parsons, your company meant more than you may know.

Finally, I thank my parents, who have loved and supported me through everything, and have always encouraged me to succeed, and my brother Jamey, who has always been there for me through all the trials and tribulations of graduate school.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

°C	degrees Celsius
µL	microliter
µmol	micromole
^{13}C	carbon-13 nucleus
^{1}H	proton
5AU	5-aminouracil
AcOH	acetic acid
Ad^{HA}	2-hydroxy-4-(adenin-9-yl)-butanoic acid
Ad^{HA}-D	2-hydroxy-4-(adenin-9-yl)-butanoic acid - aspatic acid amide
AFM	atomic force microscopy
AU	absorbance units
BA	barbituric acid
BA-GlcN	barbituric acid glucosamine glycoside
BL4C	butyrolactone-4-carboxylic acid
C18	octadecyl-functionalized silica
C18Aq	water-tolerant octadecyl-functionalized silica
CA	cyanuric acid
C-BMP	barbituric acid *C*-ribonucleotide
CD	circular dichroism
COSY	correlation spectroscopy
CyCo4	4-cyanuryl-butyric acid
CyCo6	6-cyanuryl-hexanoic acid

Cy^{HA}	2-hydroxy-4-cyanuryl-butanoic acid
Cy^{HA}-D	2-hydroxy-4-cyanuryl-butanoic acid – aspartic acid amide
D	aspartic acid *or* dextrorotary *or* deuterium
D_2O	deuterium oxide
DAP	2,6-diaminopurine
DAPHC	2-(2,6-diaminopurin-8-yl)-homocysteamine
DBU	1,8-diazabicyclo[5.4.0]undec-7-ene
DMF	*N*,*N*-dimethylformamide
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
DSS	4,4-dimethyl-4-silapentane-1-sulfonic acid
EtOAc	ethyl acetate
EtOH	ethanol
g	gram
GBL	γ-butyrolactone
GBTL	γ-butyrothiolactone
HCl	hydrochloric acid
HCN	hydrogen cyanide
HCO_2H	formic acid
HMBC	heteronuclear multiple bond correlation spectroscopy
HOMO	highest occupied molecular orbital
HPLC	high performance liquid chromatography
hr	hour
HRMS	high resolution mass spectrometry
HSQC	heteronuclear single quantum correlation

Hz	hertz
iPrOH	isopropanol
IR	infrared spectroscopy
kg	kilogram
L	liter *or* levorotary
LC/MS	liquid chromatography mass spectrometry
LUMO	lowest unoccupied molecular orbital
M	molar
m/z	mass-to-charge ratio
MAC	minimum assembly concentration
mAU	milli-absorbance units
MeCN	acetonitrile
Mel^{HA}	2-hydroxy-4-(*N*-melaminyl)-butanoic acid
Mel^{HA}-D	2-hydroxy-4-(*N*-melaminyl)-butanoic acid - aspartic acid amide
MeOH	Methanol
mg	milligram
MHz	megahertz
min	minute
mL	milliliter
mm	millimeter
mmol	millimole
MMP	melamine *N*-ribonucleotide
mol	mole
MT2C	2-methylthiazolidine-2-carboxylic acid
MVK	methyl vinyl ketone

NaCl	sodium chloride
NaOH	sodium hydroxide
NH_4HCO_2	ammonium formate
NH_4OAc	ammonium acetate
nm	nanometer
NMR	nuclear magnetic resonance spectroscopy
NMS	*N*-methylsuccinimide
NOE	nuclear Overhauser effect
PCR	polymerase chain reaction
pD	negative logarithm of deuterium ion concentration
pH	negative logarithm of hydrogen ion concentration
pK_a	negative logarithm of acid dissociation constant
ppm	parts per million
R	rectus stereochemistry
RNA	ribonucleic acid
ROE	rotating frame nuclear Overhauser effect
ROESY	rotating frame nuclear Overhauser effect spectroscopy
rt	room temperature
S	sinister stereochemistry
sec	second
T	temperature
T2C	thiazolidine-2-carboxylic acid
TAP	2,4,6-triaminopyrimidine
TAP-GalA	2,4,6-triaminopyrimidine galacturonic acid glycoside
TAP-Glc	2,4,6-triaminopyrimidine glucoside

TAP-Glc6P	2,4,6-triaminopyrimidine glucose-6-phosphate glycoside
TAP-GlcA	2,4,6-triaminopyrimidine glucuronic acid glycoside
TAP-GlcNAc	2,4,6-triaminopyrimidine *N*-acetylglucosamine glycoside
TARC	2,4,6-triaminopyrimidine β-*C*-ribofuranoside
TBD	1,5,7-triazabicyclo[4.4.0]dec-5-ene
TEA	triethylamine
TetAP	2,4,5,6,-tetraaminopyrimidine
TFA	trifluoroacetic acid
TMSCl	trimethylsilyl chloride
TNA	threose nucleic acid
TSP	trimethylsilylpropanoic acid
UV	ultraviolet spectroscopy
XNA	xenonucleic acid

SUMMARY

The origin of life on Earth is a poorly understood phenomenon. All extant life evolves by Darwinian evolution, which requires an informational, hereditary polymer, such as DNA or RNA. The emergence of nucleic acids, or primitive informational polymers resembling nucleic acids, is therefore a central theme in prebiotic chemistry. The influential RNA world hypothesis states that, at some early stage in the evolution of life, before the establishment of the genetic code, RNA was the sole (or primary) biopolymer, performing both informational and catalytic functions. In some interpretations of the RNA world hypothesis, RNA is the first polymer to arise from prebiotic chemical processes. However, the prebiotic synthesis of RNA is problematic: forming the various components of RNA selectively and covalently linking them is prebiotically difficult. For these reasons, it has been hypothesized that RNA is the product of evolution. In this view of the origin of nucleic acids, RNA is the penultimate member in an evolutionary series of nucleic acids, starting with the first informational polymer to arise on the early Earth: the proto-nucleic acid. The chemical components of the proto-nucleic acid, and of pre-RNAs, are not necessarily the same as those in RNA, but their functions were similar.

In this dissertation, I describe efforts to elucidate general physical organic principles that dictate the spontaneous prebiotic emergence of proto-nucleic acids. Candidate proto-nucleic acid components, such as noncanonical nucleobases and noncanonical backbone motifs, are described, and criteria to judge their candidacy, such as chemical reactivity and propensity for oligomerization and self-assembly in water, are discussed. Although the chemical space of informational polymers is vast, the principles

arrived at greatly reduce the size of this space by ruling out inviable chemical motifs. This work culminates with the introduction of a new class of informational polymer that is considered a strong candidate for proto-RNA. Important general principles for the evolution of proto-nucleic acids and pre-RNAs are also discussed.

CHAPTER 1. INTRODUCTION

1.1 The Origin of Life on Earth

All life on Earth stores hereditary information in the informational polymer deoxyribonucleic acid (DNA). DNA is an informational polymer that can be replicated in a template-directed manner by protein enzymes. In order to implement the instructions in DNA, it is transcribed into a related polymer, ribonucleic acid (RNA), by protein enzymes. The instructions in a messenger RNA are then translated into proteins by the ribosome, an RNA enzyme. This flow of molecular information, known as the Central Dogma of Molecular Biology [1], is evolutionarily sophisticated. Unlike common descent, which is apparent from the universality of the Central Dogma, the chemical processes through which life began are not at all obvious. Because the biosyntheses of DNA and protein are interdependent, it appears that some informational molecular system must have preceded them.

1.2 The RNA World Hypothesis

Unlike DNA and proteins, which exclusively fill information storage and functional (i.e., catalytic, structural, signaling, etc.) roles, respectively, RNA is capable of filling both roles. This is evident from the Central Dogma: messenger RNA is informational, being synthesized from a complementary DNA strand and being translated into a specific amino acid sequence, while ribosomal RNA is functional, catalyzing peptide bond formation. Because of this versatility of RNA, it has been hypothesized that an early state of evolution existed in which RNA was the sole polymer of life [2]. This so-called “RNA world

hypothesis" can be interpreted in several ways. In one popular interpretation, RNA is not only the immediate predecessor of DNA, but the first polymer of life to arise directly from prebiotic chemical processes on the early Earth [3].

1.2.1 Problems with the Spontaneous Prebiotic Emergence of RNA

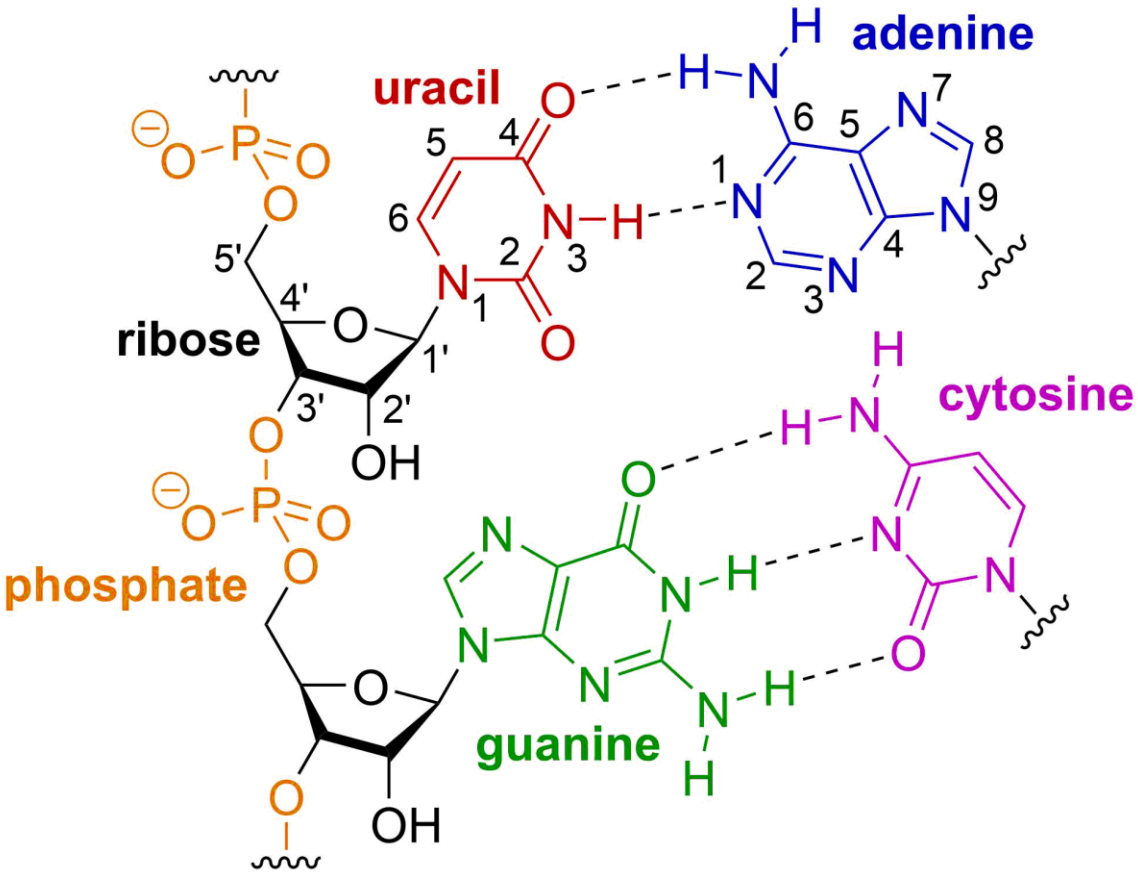


Figure 1.1. The chemical structure of ribonucleic acid (RNA). RNA is an informational polymer with a regular ribose (black) phosphate (orange) backbone and four possible nucleobase substituents: adenine (a purine, shown in blue), uracil (a pyrimidine, shown in red), guanine (a purine, shown in green), and cytosine (a pyrimidine, shown in magenta). Uracil is shown with systematic pyrimidine numbering. Adenine is shown with systematic purine numbering. The 5′ ribose is shown with systematic sugar numbering.

There are a number of problems associated with the spontaneous prebiotic emergence of RNA. First, the glycosidic bond that links the ribose moiety to a canonical

nucleobase is difficult to form under plausibly prebiotic conditions: it does not form from a direct reaction with ribose and the canonical nucleobases in water [4]. It can be formed, for example, if position 1′ (the anomeric position) of ribose is derivatized with a moiety that can serve as a better leaving group than water [5]. However, the formation of these activated forms of ribose invokes the presence of hydrolytically unstable and potentially depletable chemical activating agents.

Second, the ribose sugar present in RNA is difficult to form selectively in a plausibly prebiotic manner. Ribose is a "typical" sugar in the sense that it is a formal oligomer of formaldehyde, with its carbon atoms having a net oxidation state of zero. "Typical" sugars can be synthesized in the formose reaction, which is often invoked as their prebiotic source. However, the formose reaction is inherently non-selective, producing literally hundreds of sugar products, with no special preference for ribose [6]. Syntheses have been developed that attempt to surmount these two initial difficulties by building the ribose moiety and nucleobase moiety simultaneously [7, 8], but the prebiotic robustness of these syntheses has been questioned [9-11].

Third, the phosphodiester backbone linkage of RNA is difficult to form under plausibly prebiotic conditions [12]. The esterification of phosphate in water is kinetically and thermodynamically unfavorable. This is thermodynamically obvious, as esterification of phosphate requires the expulsion of an equivalent of water. Kinetically, esterification of phosphate typically requires phosphate to act as an electrophile; however, except under rather acidic conditions (pH < 2), phosphate is anionic, which greatly diminishes its electrophilicity by preventing the attack of a nucleophile by electrostatic repulsion [13]. A number of prebiotic chemical routes to phospho(di)esters have been proposed. Activated

forms of phosphate, such as trimetaphosphate [14, 15] and diamidophosphate [16], can act as phosphorylating agents, but the prebiotic formation of these activated species is not firmly established. Furthermore, these activated forms of phosphate, like the activated forms of ribose mentioned above, are hydrolytically unstable and potentially depletable. Phosphorylation of nucleosides such as adenosine with unactivated phosphate has been accomplished in urea-rich media, such as molten (or semi-molten) urea [17], or multi-component solvents that contain urea, such as formamide-urea mixtures [7], or ammonium formate-water-urea mixtures [18]. However, these urea-rich media would seriously interfere with the initial formation of nucleosides, as nitrogen-rich chemical species react destructively with ribose and other sugars [19], raising the question of how nucleosides could be formed and subsequently phosphorylated in a single environment (although geochemical scenarios have been proposed to circumvent this problem [20]). It should also be noted that phosphate, if present on the early Earth, is thought to have been primarily sequestered in insoluble minerals such as apatite [12]. Sources of phosphorus in lower oxidation states, such as phosphites [21], or the meteoritic mineral schreibersite [22], have been shown to produce phosphoesters when reacted with alcohols in water. However, these phosphorylation reactions emphatically require the presence of an oxidizing agent. Water has been suggested as the oxidizing agent in scheibersite-mediated phosphorylation, liberating hydrogen gas to form pentavalent phosphorus species [23].

In addition to these problems of chemical reactivity, there are a number of questions on the origin of RNA related to processes of selection. As alluded to above, the chemical reactions usually invoked for the prebiotic synthesis of sugars are non-selective, producing ribose in addition to other sugars [6]. Why, then, is only ribose (or deoxyribose) present in

extant nucleic acids, rather than any other possible sugar? Assuming that the sugar present in a nucleic acid, for reasons of chemical reactivity that will be discussed later, must be an aldose, and must be able to adopt a cyclic hemiacetal structure, there are two tetroses and four pentoses that could have served as trifunctional connectors in the backbone of a nucleic acid. Was the selection of ribose a result of chemical determinism from some hitherto unknown (or under-appreciated) selective synthesis of ribose (or ribonucleotides), or was ribose "chosen" through some process of selection (chemical or Darwinian)?

A related problem is the chemical configuration of ribose and the phosphodiester linkage in the backbone of RNA. Ribose in its native state is present in four cyclic hemiacetal forms and, to a small extent, one linear aldehydic form. Of the four possible cyclic forms, why is the beta-furanoside form the one found in extant RNA? Furthermore, why does the phosphodiester linkage of RNA present exclusively as the 3′-5′ linkage, rather than the 2′-5′ linkage (especially considering that nonenzymatic, template-directed primer extensions of RNA produce both types of linkages [24, 25])? Studies have been published that address these issues, and suggest that the β-furanosyl form of ribose is optimal for the functions of RNA [26]. This may suggest that it was arrived at through an evolutionary optimization process.

As a digression, it is interesting to note that the cis configuration of the 2′-3′ diol moiety of RNA allows for a mechanism of spontaneous strand scission (the so-called in-line cleavage mechanism) that would not be possible for other hydroxyl group orientations in other sugars. Although this degradation reaction is typically thought of as a detriment to the prebiotic viability of RNA, could it have been necessary for the recycling of RNA in early forms of life (before the evolution of nucleases), perhaps to turn over ribozymes, or

because self-replicating systems that could turn over monomers could evolve more rapidly than those that could not?

To further complicate matters, the canonical nucleobases of RNA, adenine, guanine, uracil, and cytosine, are produced prebiotically among many other related heterocyclic compounds [27], and it is not immediately obvious why only this set of four was chosen for the information-transducing functions of RNA. On the grounds of chemical stability, these bases seem superior to others [28, 29]. However, on the grounds of chemical reactivity to form nucleosides, these bases are approximately equal to some noncanonical nucleobases, but inferior to others, as will be shown.

1.3 Proto-Nucleic Acids and the Origin of RNA

Given the many problems associated with producing RNA abiotically, it seems reasonable that RNA, rather than being produced directly from prebiotic chemical processes, is the product of evolution [30]. In this view of the origin of nucleic acids, RNA is the penultimate member in an evolutionary series of informational polymers, being preceded by more primitive, but more prebiotically viable, proto-nucleic acids, and being succeeded by DNA. The predecessors of RNA, or pre-RNAs, resembled RNA in functionality, but not necessarily in chemical structure (Figure 1.2). They are expected to be self-assembling and information-transducing by Watson-Crick-like hydrogen bonding associations between heterocyclic recognition units, and they are expected to be polymeric, with a trifunctional moiety to append the recognition units to, and a (negatively) charged moiety to maintain solubility. The canonical chemical components of extant RNA (ribose, phosphate, and the canonical nucleobases, adenine, guanine, uracil, and cytosine) were not

necessarily present in pre-RNAs. In heredity-first models of the origin of life, the primordial informational polymer, or proto-nucleic acid, must have been produced from totally abiotic processes; therefore, for the reasons stated above concerning the prebiotic difficulty in forming RNA, it may be that few or none of the canonical components were present in the proto-nucleic acid.

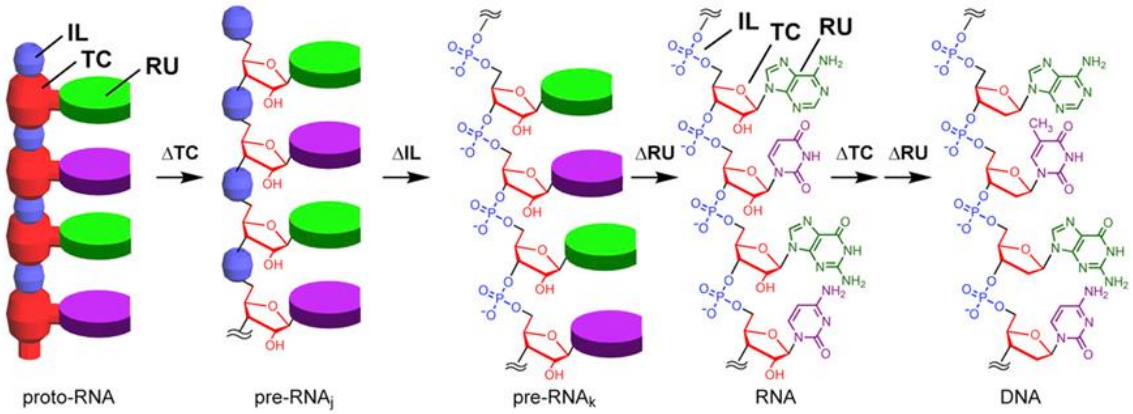


Figure 1.2. The "Grandfather's Axe" model of the origin of RNA and DNA. RNA is composed of three structural elements: the recognition unit (RU), trifunctional connector (TC), and the ionized linker (IL). More primitive nucleic acids, which may have been the ancestors of RNA, had these same structural elements, but the exact chemical identities may have been different. Sequential substitutions of these elements eventually furnished RNA and DNA. Adapted with permission from [30].

If any pre-RNAs did exist, they have left no trace in extant life. RNA itself is perhaps the oldest remaining molecular "fossil", especially with regard to the interior of the ribosome, which is conserved across all domains of life [31]. However, just because RNA is oldest chemical component of extant life, this emphatically does not mean that is was the first. Still, with no molecular evidence left from a pre-RNA world, how do we begin to decipher the chemical nature of proto-nucleic acids?

For both the recognition units and the backbone of proto-RNA, chemical reactivity is a useful parameter to start with. Recall that the canonical nucleobases of RNA

are not sufficiently reactive to form ribosides in water with unactivated ribose, and that phosphorylation is also a difficult reaction to accomplish. It is important to note, however, that chemical reactivity is not a unidimensional parameter: although sufficient reactivity is important to consider, the manner of reactivity, in terms of nucleophilicity/electrophilicity, hardness/softness, valency, and reversibility, is equally important.

Another parameter that is especially relevant to chemical evolution is the propensity for self-assembly. Self-assembly in nucleic acids is necessary for template-directed replication, but is also important for its general property of inhibiting hydrolysis [32]. This inhibition is entropic in nature: in order to undergo hydrolysis, a hydrolysable bond must sample conformational states that are close to the transition state of, for example, nucleophilic attack of water. In an assembled state, oligomers are conformationally restricted, and are therefore less likely to occupy states close to the hydrolysis transition state. In an oligomerization system using reversible bonds, chemical evolution can favor the formation of self-assembling oligomers by recycling monomers of oligomers that do not assemble.

The process of self-assembly of extant nucleic acids is facilitated by Watson-Crick base pairing between the canonical nucleobases, and emerges only on the oligomer level; i.e., mononucleotides of the canonical bases do not spontaneously self-assemble [33](with the exception of guanosine in G tetrads). However, this limited extent of self-assembly is not necessarily the case for nucleosides or nucleotides of certain noncanonical nucleobases [34, 35].

Self-assembly of a candidate proto-nucleic acid must also be permitted by the geometry of the backbone. Therefore, even if a candidate monomer is capable of self-assembly, it may be disqualified on structural grounds of its corresponding polymer. Candidate backbones can also be evaluated by their repeat unit lengths. DNA, RNA, and a number of synthetic nucleic acid analogues, such as PNA, all have 6-atom (or 6-bond) repeats [36]. Threose nucleic acid (TNA) [37] and glycol nucleic acid (GNA) [38] successfully undergo self-assembly with a 5-atom repeat. 2′-5′ RNA, an isomer of natural 3′-5′ RNA, forms duplexes with a 7-atom repeat, albeit with much lower thermal stability than 3′-5′ RNA [39]. Polymeric sequences that are shorter or longer than these may be considered less likely to successfully form supramolecular assemblies.

1.3.1 Noncanonical Nucleobases

Given that many noncanonical nitrogenous heterocyclic compounds capable of molecular recognition and self-assembly [35, 40-42] are produced in model prebiotic reactions [43] and detected in meteorites [44], there is no strong reason to assume that the earliest genetic polymers necessarily contained only the canonical bases. In addition to pyrimidines and purines, several other heterocyclic classes of compounds may be prebiotically relevant, including pteridines [45], triazines [43], pyrazines [46], imidazoles (including hydantoins) [47], and triazoles [48]. Although this chemical space is vast, the number of viable candidates can be reduced based on certain chemical principles. As mentioned above, appropriate chemical reactivity and propensity for self-assembly are important criteria (both of which are affected by the choice of backbone, as differences in reactivity may be necessary to append nucleobases to a candidate backbone, and as differences in backbone geometry may favor or disfavor different modes of self-assembly).

Another simple but important criterion is aqueous solubility. Although large heterocyclic compounds, such as pteridines, might assemble more efficiently than smaller ones due to a larger hydrophobic stacking surface, this same chemical properties greatly reduces their solubility in water, preventing them from participating in monomer-forming reactions.

Assuming that the proto-nucleic acid, or some pre-RNA, was composed of glycosides (as a generalization of ribonucleosides), then one of the most pertinent criteria for the viability of a candidate ancestral nucleobase is the ability to react with ribose and other sugars in water without the use of activated intermediates. This restricts the space of nucleobases to those that contain nucleophilic exocyclic amino groups, protic methylene groups, or electron-rich methine groups, as only these functionalities can react with sugars in water (for reasons that will be elaborated on in later sections). However, if the candidate ancestral backbone is formed from a different type of electrophilic trifunctional connector, this constraint can be modified or relaxed.

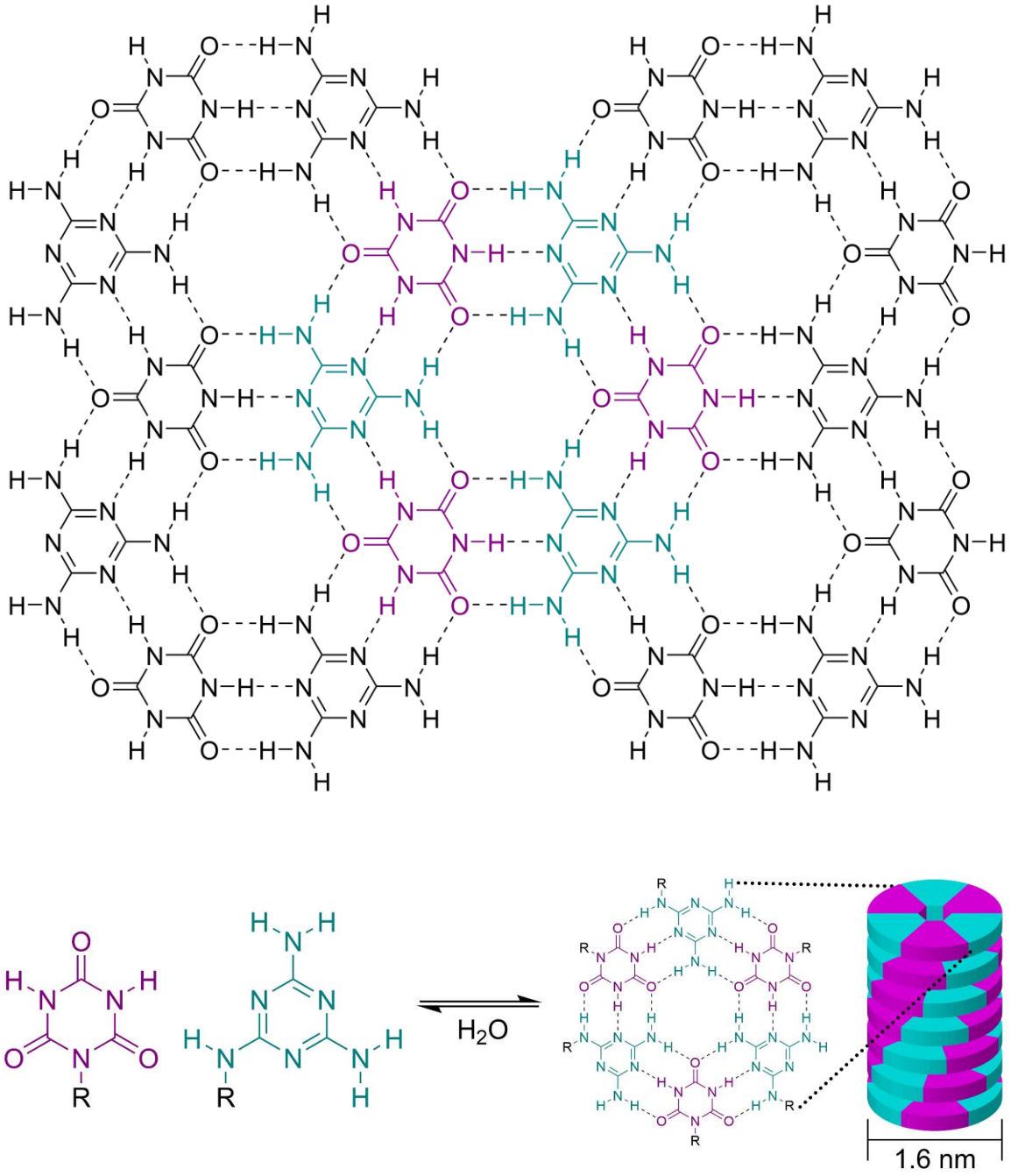


Figure 1.3. The melamine-cyanuric acid supramolecular assembly system. Top: The melamine-cyanuric acid crystal lattice, with a hexad highlighted in cyan (melamine) and magenta (cyanuric acid). Bottom: Monosubstituted derivatives of melamine and cyanuric acid spontaneously self-assemble in water.

Self-assembly by complementary hydrogen bonding of nitrogenous heterocycles can give complexes with different cardinalities (i.e., whether the complex is a duplex, tetraplex, hexaplex, etc.). Although extant nucleic acids typically form duplexes (with triplexes and tetraplexes occasionally observed), this does not necessarily mean that proto-nucleic acids or pre-RNAs must also have semi-exclusively formed duplexes. In fact, due to the larger hydrophobic stacking surface area associated with larger complexes (such as hexaplexes), these higher-cardinality complexes form at much lower monomer concentrations than duplexes [49]. The most commonly encountered hexaplex motifs are based off of the melamine-cyanuric acid crystal lattice, which is formed from the hexad substructure containing three melamine units and three cyanuric acid units arranged according to complementary hydrogen bonding (Figure 1.3) [50, 51]. If one or both of these triazines is derivatized on one hydrogen bonding face, then isolated hexads can be formed in solution [50, 51]. If the derivitization confers a peripheral charge, then hexad assemblies of this type can be formed in water [34, 49]. The hexad, once assembled, presents a very large hydrophobic stacking surface – approximately 1.6 nm^2 – to the aqueous solvent. To sequester this large hydrophobic surface, the individual hexads stack on each other, forming noncovalent fibers that can be thousands of hexads in length. In addition to the triazines melamine and cyanuric acid, hexads with the same hydrogen-bondning motif can be formed from the pyrimidines barbituric acid [35] and 2,4,6-triaminopyrimidine (TAP) [34, 49]. With the exception of TAP, all of these compounds have been formed in model prebiotic reactions [43]. Furthermore, these compounds possess the appropriate chemical moieties to react with sugars. Melamine and TAP possess exocyclic amino groups that could potentially condense with aldehydes of aldose sugars. TAP also possesses an

electron-rich, nucleophilic methine group. Barbituric acid possesses a protic methylene group that can easily be deprotonated to give the nucleophilic barbiturate anion. Cyanuric acid, unlike the other nucleobases, does not possess the proper chemical moieties to react with aldehydes; however, anticipating that it possesses the appropriate mode of reactivity to form alternative proto-nucleic acid monomers, it will remain in consideration. For these reasons, this set of four nucleobases (Figure 1.4), and close relatives thereof, will serve as our candidates for the primordial recognition units of proto-nucleic acids.

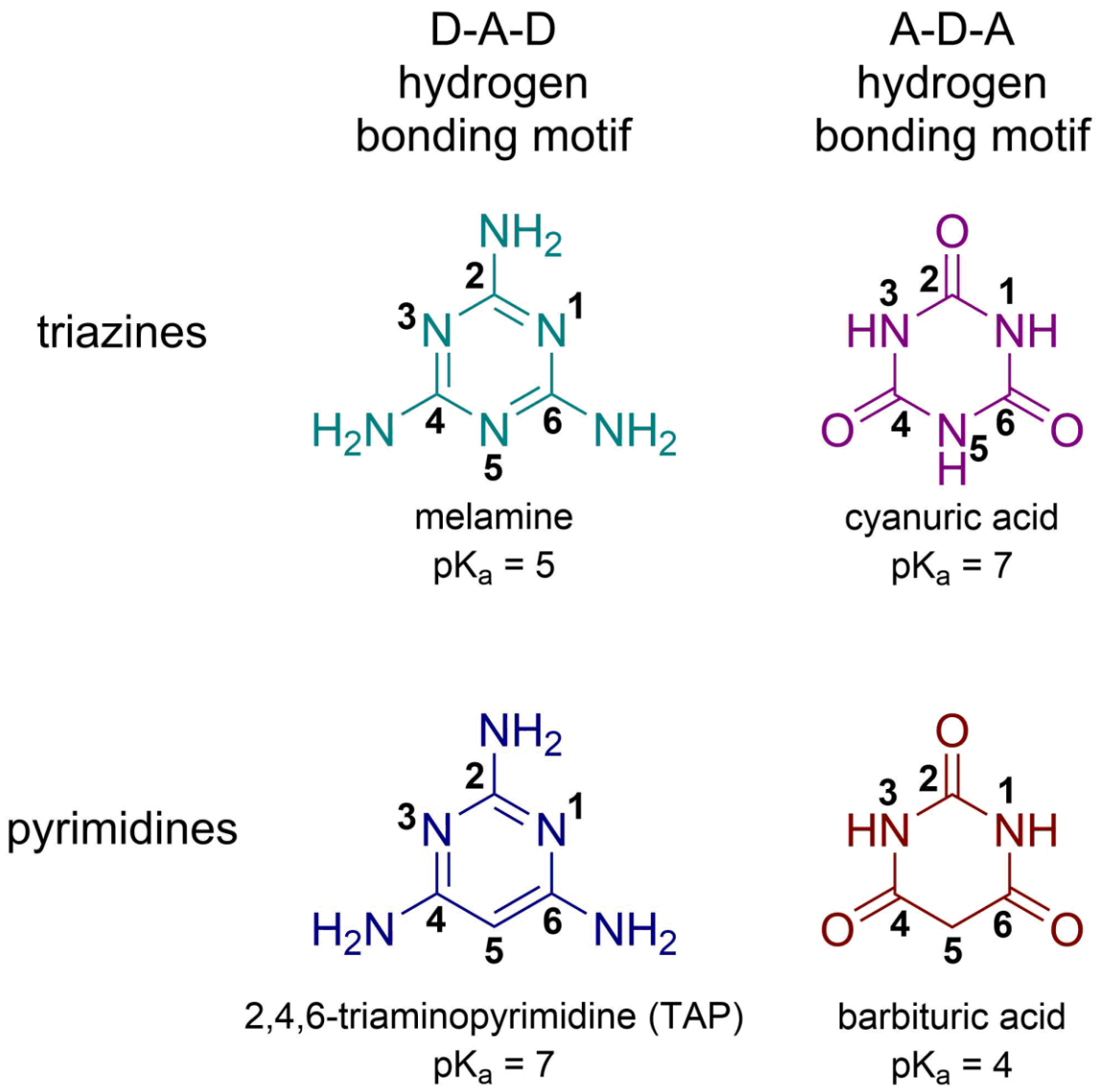


Figure 1.4. The hexad-forming noncanonical nucleobases, annotated with systematic numbering and pK$_a$ values (of conjugate acids for TAP and melamine).

1.3.2 Noncanonical Backbones

Because of the prebiotic difficulty in forming mono- and diesters of phosphate, and because of the inability to form ribose selectively, it is worthwhile to consider alternatives to the canonical RNA backbone for the earliest genetic polymers. Like the proto-nucleobases, the chemical space of candidate proto-nucleic acid backbones is vast. In order to reduce the size of this chemical space, two approaches can be taken, one of which is reminiscent of a top-down approach, and the other of a bottom-up approach.

In a top-down manner, it can be inferred from the extant nucleic acids that, at some stage in chemical evolution, sugars were selected as the trifunctional connector. It has been suggested that threose, a "simpler" sugar in the sense that it contains one fewer carbon atom than ribose, may have served as an early trifunctional connector to form threose nucleic acid (TNA) [52]. This hypothesis has some merit in that TNA forms stable duplexes, but does not truly address the problem of sugar selection, since it still leaves the question of the mechanism of selection unanswered: even though threose is "simpler" than ribose, it is still not produced selectively by model prebiotic reactions. The evaluation of the reactions of noncanonical nucleobases with a large set of sugars may provide some insight into the selection mechanism, as it may have been that selection occurred by the propensity to form stable glycosides (i.e., noncanonical nucleosides). Still, this would not satisfactorily address the problem of backbone formation, since "typical" sugars possess only one electrophilic moiety, to which the nucleobase must be appended. Therefore, in order to oligomerize a noncanonical nucleoside, a di-electrophilic moiety must be installed. Of course, if a sugar derivative is used, such as a uronic acid, then oligomeric species may be able to form.

Assuming that the original chemical bond that formed the backbone of the proto-nucleic acid was not an ester of phosphate, what chemical bonds could have taken its place? This question can be attacked with a bottom-up thought experiment. Consider a set of plausibly prebiotic electrophilic moieties, and a set of plausibly prebiotic nucleophilic moieties. From these two sets, a matrix of chemical bonds can be constructed, and each bond can be evaluated individually (Figure 1.5). To constrain this chemical space further, we will consider only bonds that form by condensation-dehydration; i.e., bonds that form with the concomitant liberation of an equivalent of water. These bonds are convenient to consider because they allow for monomer recycling by hydrolysis (see § 1.2.1 and § 1.3).

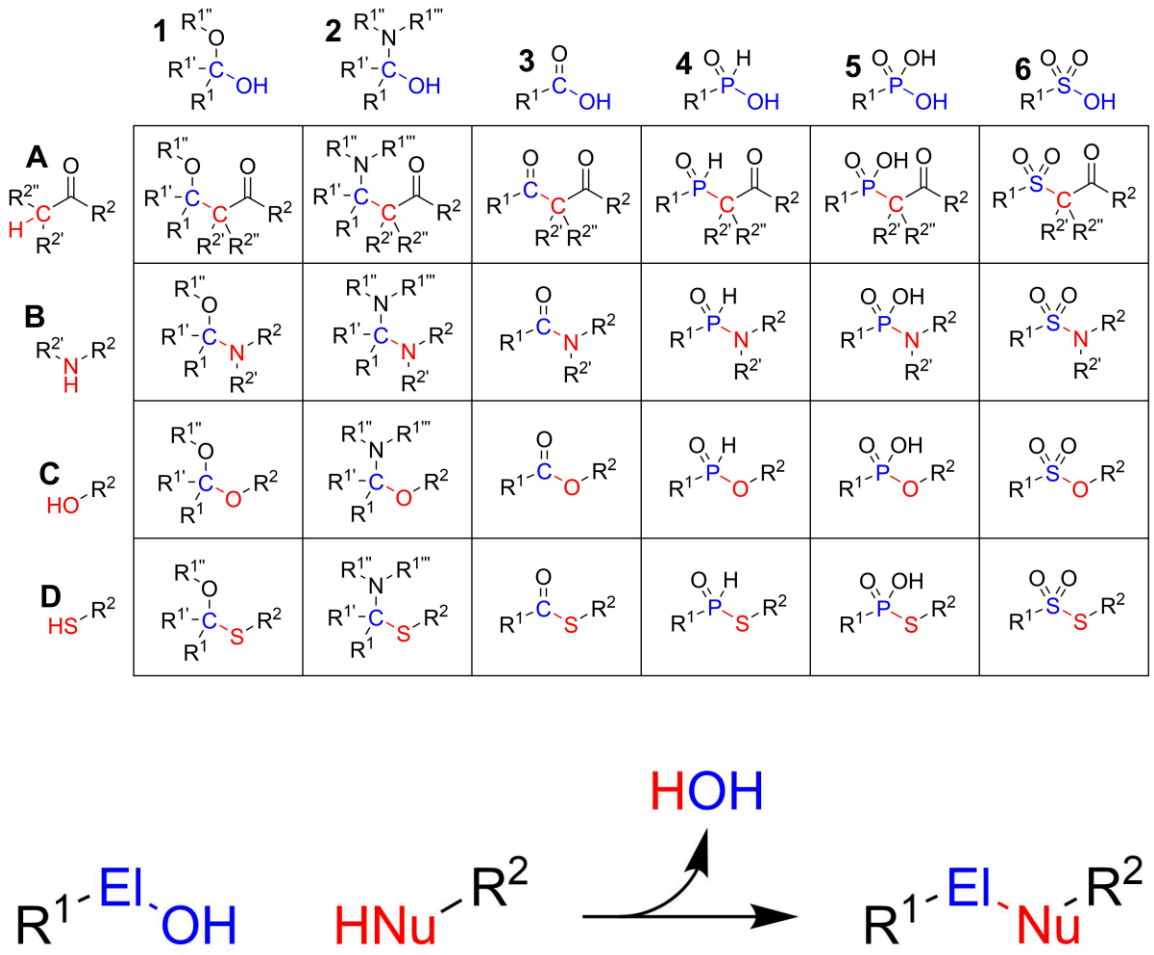


Figure 1.5. Matrix of candidate backbone linkages formed from prebiotic nucleophiles (left axis) and prebiotic electrophiles (top axis). Nucleophilic atoms and their protic hydrogen atoms are shown in red, and electrophilic atoms and their hydroxide leaving groups are shown in blue. The formal dehydration reaction of a general nucleophile with a general electrophile is shown below the matrix.

To emphasize that the linkage moieties considered are condensation-dehydration linkages, the sets of prebiotic nucleophiles and electrophiles are shown in formal structures, with nucleophilic atoms and their protic hydrogen atoms highlighted in red, and with electrophilic atoms and their formal hydroxide leaving groups highlighted in blue. These forms do not necessarily represent the exact nucleophilic or electrophilic reaction intermediates. Furthermore, some of these linkages might be formed from the fusion of two nucleophilic moieties with one di-electrophilic moiety; e.g., the phosphodiester or the

acetal, but in the set of formal electrophiles, are represented with one nucleophilic element already appended to the di-electrophilic element. Finally, this matrix is not exhaustive; it is intended only to demonstrate that a systematic survey and elimination of candidates can be performed.

From this matrix of possible backbone linkages, we seek one that can be formed with a low kinetic barrier under certain prebiotic conditions, but can persist for long enough, perhaps under some different set of conditions, to support a self-assembling system. In this regard, the acetal/ketal (Figure 1.5, row C, column 1) has been considered because of its ease of formation and cleavage under acidic conditions and its kinetic resistance to hydrolysis under basic conditions [53]. Glyoxylic acid (CHO-CO2H) has been shown to form acetal-linked thymidine dimers under drying conditions with a magnesium Lewis acid [54]. However, a yield of only 1% was observed for all dimer isomers combined. Although this does not disqualify glyoxylate acetals as ionized linkers in pre-RNAs, there are perhaps other bond types that could form more easily under prebiotic conditions to form the first nucleic acid.

In the context of proto-polypeptide formation, it has been recognized that, along with α-amino acids, α-hydroxy acids are also produced in model prebiotic reactions [55], and by a similar mechanism [55, 56]. Malic acid, the hydroxy acid analogue of aspartic acid, was found to readily oligomerize by esterification from a drying aqueous solution at acidic pH [57]. Furthermore, it was recently shown that mixtures of amino acids and hydroxy acids, when dried from an aqueous solution at acidic pH, can form depsipeptide oligomers, containing both amide (Figure 1.5, row B, column 3) and ester (Figure 1.5, row C, column 3) bonds [58]. The mechanism of depsipeptide formation seems to be initiated

by acidic-catalyzed esterification (i.e., Fischer esterification) in the dry state, followed by ester aminolysis to form amides. Because of this ease of formation of esters from a plausibly prebiotic hydration-dehydration cycle, the ester bond (and perhaps the amide bond as well) presents itself as a strong candidate for the backbone linkage present in the earliest genetic polymers.

1.4 Outlook

In this dissertation, I will describe efforts to elucidate general physical organic principles of the spontaneous emergence of self-assembling, informational oligomers (i.e., proto-nucleic acids) that could have plausibly occurred on the early Earth. Previous work has established the strong candidacy of certain noncanonical nucleobases as components of ancestral pre-RNAs, primarily on the grounds of propensity for self-assembly in water [27]. The research described herein on proto-nucleobases is primarily concerned with the chemical principles dictating their reactivity with candidate proto-RNA trifunctional connectors. Depending on the choice of trifunctional connector moiety (or choice of electrophilic moiety in general), the suitability of a given candidate nucleobase varies. Consequences for self-assembly are also observed. Chapters 2, 3, and 4 describe these investigations, and the implications for the emergence of proto-RNA and the evolution of pre-RNAs are also discussed.

A number of studies have been also conducted to identify proto-RNA backbone candidates. These studies also revealed important principles in chemical reactivity and in the propensity for self-assembly. Some candidates are ruled out, while others are found to be most suitable for intermediate pre-RNAs. Ultimately, a new class of prebiotic

compound, depsipeptide nucleic acid, is found to be a strong candidate for the proto-nucleic acid. Chapters 5 and 6 describe these investigations.

The thesis concludes with a broad discussion of general physical organic principles that dictate the formation of proto-nucleic acids. Additionally, comments are made on the possible modes of early evolution of proto-nucleic acids that are consistent with the chemical principles elucidated from this body of research.

1.5 References

[1] Crick, F., Central Dogma of Molecular Biology. Nature, 1970. 226: p. 561-563.

[2] Gesteland, R.F., T.R. Cech, and J.F. Atkins, The RNA World. 3rd ed. 2006, Cold Spring harbor, NY: Cold Spring Harbor Laboratory Press. 768.

[3] Joyce, G.F. and L.E. Orgel, Prospects for understanding the origin of the RNA world, in The RNA World, Second Edition: The Nature of Modern RNA Suggests a Prebiotic RNA World, R.F. Gesteland and J.F. Atkins, Editors. 1999, Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY. p. 49-77.

[4] Fuller, W.D., R.A. Sanchez, and L.E. Orgel, Studies in prebiotic synthesis. VI. synthesis of purine nucleosides. J. Mol. Biol., 1972. 67(1): p. 25-33.

[5] Kim, H.J. and S.A. Benner, Prebiotic stereoselective synthesis of purine and noncanonical pyrimidine nucleotide from nucleobases and phosphorylated carbohydrates. Proc. Natl. Acad. Sci. U S A, 2017. 114(43): p. 11315-11320.

[6] Decker, P., P. Schweer, and R. Pohlmann, Identification of formose sugars, presumable prebiotic metabolites, using capillary gas chromatography/gas chromatography-mass spectroscopy of n-butoxime trifluoroacetates on OV-225. J. Chromatogr., 1982. 244: p. 281-291.

[7] Powner, M.W., B. Gerland, and J.D. Sutherland, Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. Nature, 2009. 459(7244): p. 239-242.

[8] Stairs, S., et al., Divergent prebiotic synthesis of pyrimidine and 8-oxo-purine ribonucleotides. Nature Communications, 2017. 8: p. 12.

[9] Shapiro, R., A simpler origin for life. Sci Am, 2007. 296(6): p. 46-53.

[10] Eschenmoser, A., Etiology of potentially primordial biomolecular structures: From vitamin B12 to the nucleic acids and an inquiry into the chemistry of life's origin: A retrospective. Angewandte Chemie-International Edition, 2011. 50(52): p. 12412-12472.

[11] Benner, S.A., H.-J. Kim, and M.A. Carrigan, Asphalt, water, and the prebiotic synthesis of ribose, ribonucleosides, and RNA. Acc. Chem. Res., 2012. 45: p. 2025-2034.

[12] Schwartz, A.W., Phosphorus in prebiotic chemistry. Philosophical transactions of the Royal Society of London, Series B: Biological Sciences, 2006: p. 1743-1749.

[13] Westheimer, F.H., Why Nature chose phosphates. Science, 1987. 235: p. 1173-1178.

[14] Saffhill, R., Selective phosphorylation of the cis-2',3'-diol of unprotected ribonucleosides with trimetaphosphate in aqueous solution. J. Org. Chem., 1970. 35(9): p. 2881-3.

[15] Keefe, A.D. and S.L. Miller, Potentially prebiotic syntheses of condensed phosphates. Origins of Life and Evolution of the Biosphere, 1996. 26(1): p. 15-25.

[16] Gibard, C., et al., Phosphorylation, oligomerization and self-assembly in water under potential prebiotic conditions. Nat. Chem., 2018. 10(2): p. 212-217.

[17] Lohrmann, R. and L.E. Orgel, Urea-inorganic phosphate mixtures as prebiotic phosphorylating agents. Science, 1971. 171(3970): p. 490-494.

[18] Burcar, B., et al., Darwin's Warm Little Pond: A one-pot reaction for prebiotic phosphorylation and the mobilization of phosphate from minerals in a urea-based solvent. Angew. Chem. Int. Ed. Engl., 2016. 55: p. 13249-13253.

[19] Weber, A.L., The sugar model: Autocatalytic activity of the triose-ammonia reaction. Origins of Life and Evolution of Biospheres, 2007. 37(2): p. 105-111.

[20] Patel, B.H., et al., Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. Nat. Chem., 2015. 7: p. 301-307.

[21] De Graaf, R.M. and A.W. Schwartz, Thermal synthesis of nucleoside H-phosphonates under mild conditions. Origins of Life and Evolution of the Biosphere, 2005. 35(1): p. 1-10.

[22] Gull, M., et al., Nucleoside phosphorylation by the mineral schreibersite. Scientific Reports, 2015. 5: p. 6.

[23] Pasek, M.A., J.P. Dworkin, and D.S. Lauretta, A radical pathway for organic phosphorylation during schreibersite corrosion with implications for the origin of life. Geochimica Et Cosmochimica Acta, 2007. 71(7): p. 1721-1736.

[24] Kanavarioti, A., L.F. Lee, and S. Gangopadhyay, Relative reactivity of ribosyl 2 '-OH vs. 3 '-OH in concentrated aqueous solutions of phosphoimidazolide activated nucleotides. Origins of Life and Evolution of the Biosphere, 1999. 29(5): p. 473-487.

[25] Engelhart, A.E., M.W. Powner, and J.W. Szostak, Functional RNAs exhibit tolerance for non-heritable 2 '-5 ' versus 3 '-5 ' backbone heterogeneity. Nature Chemistry, 2013. 5(5): p. 390-394.

[26] Eschenmoser, A., Chemical etiology of nucleic acid structure. Science, 1999. 284(5423): p. 2118-2124.

[27] Cafferty, B.J. and N.V. Hud, Was a pyrimidine-pyrimidine base pair the ancestor of Watson-Crick base pairs? Insights from a systematic approach to the origin of RNA. Israel Journal of Chemistry, 2015. 55(8): p. 891-905.

[28] Rios, A.C. and Y. Tor, On the Origin of the Canonical Nucleobases: An Assessment of Selection Pressures across Chemical and Early Biological Evolution. Israel Journal of Chemistry, 2013. 53(6-7): p. 469-483.

[29] Rios, A.C., H.T. Yu, and Y. Tor, Hydrolytic fitness of N-glycosyl bonds: comparing the deglycosylation kinetics of modified, alternative, and native nucleosides. Journal of Physical Organic Chemistry, 2014: p. 173-180.

[30] Hud, N.V., et al., The origin of RNA and 'My Grandfather's Axe'. Chem. Biol., 2013. 20: p. 466-474.

[31] Petrov, A.S., et al., Evolution of the ribosome at atomic resolution. Proceedings of the National Academy of Sciences of the United States of America, 2014. 111(28): p. 10251-10256.

[32] Runnels, C.M., et al., Folding, Assembly, and Persistence: The Essential Nature and Origins of Biopolymers. Journal of Molecular Evolution, 2018. 86(9): p. 598-610.

[33] Ts'o, P.O.P., Bases, nucleosides, and nucleotides, in Basic Principles in Nucleic Acid Chemistry, P.O.P. Ts'o, Editor. 1974, New York: Academic Press. p. 453-584.

[34] Chen, M.C., et al., Spontaneous prebiotic formation of a β-ribofuranoside that self-assembles with a complementary heterocycle. Journal of the American Chemical Society, 2014. 136(15): p. 5640-5646.

[35] Cafferty, B.J., et al., Spontaneous formation and base pairing of plausible prebiotic nucleotides in water. Nature Communications, 2016. 7: p. 11328.

[36] Nielsen, P.E., et al., Sequence-Selective Recognition of DNA By Strand Displacement With a Thymine-Substituted Polyamide. Science, 1991. 254(5037): p. 1497-1500.

[37] Schöning, K.U., et al., Chemical etiology of nucleic acid structure: The alpha-threofuranosyl-(3'->2') oligonucleotide system. Science, 2000. 290: p. 1347-1351.

[38] Zhang, L.L., A. Peritz, and E. Meggers, A simple glycol nucleic acid. Journal of the American Chemical Society, 2005. 127(12): p. 4174-4175.

[39] Sawai, H., J. Seki, and H. Ozaki, Comparative studies of duplex and triplex formation of 2'-5' and 3'-5' linked oligoribonucleotides. Journal of Biomolecular Structure & Dynamics, 1996. 13(6): p. 1043-1051.

[40] Li, C., et al., Formation of supramolecular assemblies and liquid crystals by purine nucleobases and cyanuric acid in water: implications for the possible origins of RNA. Physical Chemistry Chemical Physics, 2016. 18(30): p. 20091-20096.

[41] Avakyan, N., et al., Reprogramming the assembly of unmodified DNA with a small molecule. Nature Chemistry, 2016. 8(4): p. 368-376.

[42] Yokosawa, T., et al., A step into the RNA world: Conditional analysis of hydrogel formation of adenosine 5'-monophosphate induced by cyanuric acid. Biosystems, 2017. 162: p. 53-58.

[43] Menor-Salvan, C., et al., Synthesis of pyrimidines and triazines in ice: Implications for the prebiotic chemistry of nucleobases. Chemistry-a European Journal, 2009. 15(17): p. 4411-4418.

[44] Callahan, M.P., et al., Carbonaceous meteorites contain a wide range of extraterrestrial nucleobases. Proceedings of the National Academy of Sciences of the United States of America, 2011. 108(34): p. 13995-13998.

[45] Marin-Yaseli, M.R., C. Mompean, and M. Ruiz-Bermejo, A Prebiotic Synthesis of Pterins. Chemistry, 2015. 21(39): p. 13531-4.

[46] Weber, A.L., Sugar-driven prebiotic synthesis of 3,5(6)-dimethylpyrazin-2-one: A possible nucleobase of a primitive replication process. Origins of Life and Evolution of Biospheres, 2008. 38(4): p. 279-292.

[47] Menor-Salvan, C. and M.R. Marin-Yaseli, A new route for the prebiotic synthesis of nucleobases and hydantoins in water/ice solutions involving the photochemistry of acetylene. Chemistry-a European Journal, 2013. 19(20): p. 6488-6497.

[48] Kolb, V.M., J.P. Dworkin, and S.L. Miller, Alternative bases in the RNA world: The prebiotic synthesis of urazole and its ribosides. J. Mol. Evol., 1994. 38: p. 549-557.

[49] Cafferty, B.J., et al., Efficient self-assembly in water of long noncovalent polymers by nucleobase analogues. Journal of the American Chemical Society, 2013. 135(7): p. 2447-2450.

[50] Seto, C.T. and G.M. Whitesides, Self-assembly based on the cyanuric acid melamine lattice. Journal of the American Chemical Society, 1990. 112(17): p. 6409-6411.

[51] Seto, C.T. and G.M. Whitesides, Molecular self-assembly through hydrogen-bonding - supramolecular aggregates based on the cyanuric acid.Melamine lattice. Journal of the American Chemical Society, 1993. 115(3): p. 905-916.

[52] Orgel, L., A simpler nucleic acid. Science, 2000. 290: p. 1306-1307.

[53] Hud, N.V. and F.A.L. Anet, Intercalation-mediated synthesis and replication: a new approach to the origin of life. Journal of Theoretical Biology, 2000. 205(4): p. 543-562.

[54] Bean, H.D., et al., *Glyoxylate as a backbone linkage for a prebiotic ancestor of RNA.* Origins of Life and Evolution of the Biosphere, 2006. **36**(1): p. 39-63.

[55] Miller, S.L., Production of some organic compounds under possible primitive earth conditions. Journal of the American Chemical Society, 1955. 77: p. 2351-2361.

[56] Miller, S.L., The mechanism of synthesis of amino acids by electric discharges. Biochimica Et Biophysica Acta, 1957. 23(3): p. 480-489.

[57] Mamajanov, I., et al., Ester formation and hydrolysis during wet-dry cycles: generation of far-from-equilibrium polymers in a model prebiotic reaction. Macromolecules, 2014. 47: p. 1334-1343.

[58] Forsythe, J.G., et al., Ester-Mediated Amide Bond Formation Driven by Wet-Dry Cycles: A Possible Path to Polypeptides on the Prebiotic Earth. Angewandte Chemie-International Edition, 2015. 54(34): p. 9871-9875.

CHAPTER 2. THE AQUEOUS SYNTHESIS AND SELF-ASSEMBLY OF NONCANONICAL NUCLEOSIDES AND NUCLEOTIDES

2.1 Introduction

The study of noncanonical nucleosides in a prebiotic context began in 1971 with Orgel's synthesis of inosine from ribose and hypoxanthine in the presence of magnesium chloride and sodium trimetaphosphate in the dry state at 100°C [1, 2]. Although the authors state that the formation of phosphorylated derivatives does not seem to be necessary, no inosine was produced in the absence of trimetaphosphate. The first deliberate prebiotic synthesis of a noncanonical nucleoside in water that occurred explicitly without the use of a chemical activating agent was of urazole ribosides by Kolb and Miller in 1994 [3]. The reaction of urazole (1,2,4-triazolidine-3,5-dione) with ribose produced the α- and β-pyranosides and furanosides, with the β-pyranoside produced in the greatest yield. Although not for explicitly prebiotic research purposes, glycosides of barbituric acid (pyrimidine-2,4,6(1H,3H,5H)-trione, or 6-hydroxyuracil) were produced from the aqueous reaction of barbituric acid with glucose and other hexoses by Gonzalez and coworkers in 1986 [4].

The conspicuous ability of certain potentially prebiotic noncanonical nucleobases to react with ribose and other sugars (where the canonical nucleobases fail to react) supports the hypothetical existence of nucleic acid ancestors of RNA that were more easily formed from prebiotic processes, or in early stages of the evolution of life. This hypothesis was further advanced by Hud and coworkers in 2013 when it was a found that another

noncanonical nucleobase, 2,4,6-triaminopyrimidine (TAP), was able to react with ribose without chemical activation to produce a number of riboside products, with the β-*C*-ribofuranoside (designated TARC) produced in the greatest yield [5]. It was found that TARC, in the presence of borate, could spontaneously self-assemble in water with a complementary heterocycle, cyanuric acid, to form micron-length stacked-hexad supramolecular assemblies.

In this chapter, syntheses are described for various glycosides of barbituric acid, melamine, and TAP. The mechanism of glycosylation is discussed. The self-assembly properties of some glycosides are described.

2.2 Experimental Procedures

2.2.1 Materials

Melamine, barbituric acid, 2,4,6-triaminpyrimidine, and cyanuric acid were purchased from Acros Organic. D-Ribose-5-phosphate disodium salt, D-ribose, D-glucose, D-glucosamine hydrochloride, D-glucose-6-phosphate disodium salt, D-glucuronic acid, and D-galacturonic acid were purchased from Sigma Aldrich. All compounds were used as received.

2.2.2 Sample Preparation

Unless otherwise noted, supramolecular assemblies were formed by combining the pairing compounds in water, either in the form of the parent heterocycles, corresponding glycosides, or both forms, at 50 mM in each heterocycle total. Because of the high yielding synthesis of the *C*-BMP (between 80-95% yield), the crude reaction mixture was used for

spectroscopic analysis without purification. Solutions were pH-adjusted by the addition of NaOH or HCl. All solutions gelled within minutes upon mixing complementary molecules and adjusting pH, except for the assemblies formed from MMP and barbituric acid. Experiments evaluating the assemblies of MMP and barbituric acid involved first incubating the assemblies at 5°C overnight at pH 5. After this time, a hydrogel was present.

2.2.3 Analytical Techniques

Spectroscopic analysis of the supramolecular assemblies was performed using circular dichroism (CD) and ^{1}H NMR. CD analysis was carried out on a Jasco J-720 CD spectrometer equipped with a six cell Quantum Northwest Peltier temperature controller. Strain free 0.01 mm demountable cells from Starna were used. NMR spectra were collected on a Bruker DRX 500 MHz NMR spectrometer and were the sum of 32 transients. All molecules were D_2O-exchanged and lyophilized prior to analysis in D_2O with an internal standard of TSP at 1.11 mM.

UV-LC/MS analysis of glycosylation reactions was performed with a 3.5 μm XBridge amide column running a linear gradient of 90% MeCN/10% NH_4OAc buffer 10 mM pH 9 to 60% MeCN/40% NH_4OAc buffer 10 mM pH 9 over a period of 7 minutes with a flow rate of 0.5 mL/min.

2.2.4 Glycoside Synthesis

***C*-BMP:** Barbituric acid (2.5 mmol) and ribose-5-phosphate (2.5 mmol) were dissolved in 5 mL of H_2O and the pH was adjusted to 9.0 with NaOH. The solution was stirred for 24 hours at 20°C. This solution was then loaded onto a gravity column containing QAE

Sephadex A-25 anion exchange media and eluted with a gradient of NH_4HCO_3 buffer from 50 mM to 0.5 M. The fractions containing product were combined. The maximum yield of *C*-BMP was 90% (α-*C*-BMP 26%, β-*C*-BMP 64%). HRMS (neg. m/z): $C_9H_{12}N_2O_{10}P^-$ theoretical mass: 339.0235, actual mass: 339.0243. α-*C*-BMP: ^{1}H NMR (500 MHz, D_2O) δ 4.94 (d, 4.2 Hz, H1'); 3.85 (dd, 4.2, X Hz, H2'); 3.93 (dd, 4.1, 3.5 Hz, H3'); 3.70 (m, H4'); 3.59 (m, H5'a); 3.46 (m, H5'b). ^{13}C NMR (126 MHz, D2O) δ 167.1 (C4/C6); 152.7 (C2); 85.5 (C5); 79.5 (d, 8.2 Hz, C4'); 75.1 (C1'); 74.3 (C2'); 72.7 (C3'); 63.6 (d, 4.4, C5'). β-*C*-BMP: ^{1}H NMR (500 MHz, D_2O) δ 4.51 (d, 5.7 Hz, H1'); 4.29 (t, 5.7 Hz, H2'); 3.86 (t, 6.2 Hz, H3'); 3.54 (m, H4'); 3.58 (m, H5'a); 3.46 (m, H5'b). ^{13}C NMR (126 MHz, D_2O) δ 166.4 (C4/C6); 153.1 (C2); 85.7 (C5); 80.8 (d, 7.9 Hz, C4′); 78.6 (C1'); 70.9 (C2'); 70.5 (C3'); 63.8 (d, 4.6 Hz C5').

MMP: Melamine (1 mmol) and ribose-5-phosphate (1 mmol) were dissolved in 5 mL of H_2O and the pH was adjusted to 5.0 with HCl. The solution was stirred at 65°C for 24 hours. This solution was then loaded onto a gravity column containing SP Sephadex C-50 cation exchange media and eluted with NH_4OAc buffer, 50 mM, pH 4.31. The fractions containing product were lyophilized, redissolved in water, and pooled. The maximum yield of MMP was 55% (α-MMP 26.4%, β-MMP 28.6%). HRMS (neg. m/z): $C_8H_{14}N_6O_7P^-$ theoretical mass: 337.0667, actual mass: 337.0659. α-MMP: ^{1}H NMR (500 MHz, D_2O) δ 5.46 (d, 4.1 Hz, H1'); 3.97 (dd, 4.7, 1.9 Hz, H3'); 3.94(dd, 4.7, 4.1 H2'); 3.74 (m, H4'); 3.64 (m, H5'a); 3.55 (m, H5'b). ^{13}C NMR (126 MHz, D_2O) δ 164.4 (C2); 164.2 (C3/C4); 81.0 (C1'); 79.5 (d, 8.2 Hz, C4'); 70.0 (C3'); 69.9 (C2'); 63.4 (d, 5.3, C5'). β-MMP: ^{1}H NMR (500 MHz, D_2O) δ 5.25 (d, 6.6, H1'); 3.89 (dd, 5.4, 3.2 Hz, H3'); 3.85 (dd, 6.6, 5.4 Hz,

H2'); 3.75 (m, H4'); 3.49 (m, 3.54, H5'a,b). ^{13}C NMR (126 MHz, D_2O) δ 165.1 (C2); 164.9 (C3/C4); 84.2 (C1'); 82.3 (d, 8.0 Hz, C4'); 73.1 (C2'); 70.5 (C3'); 64.2 (d, 4.5 Hz, C5').

All other compounds were synthesized as described.

2.3 Formation and Supramolecular Assembly of Complementary Noncanonical Nucleotides in Water

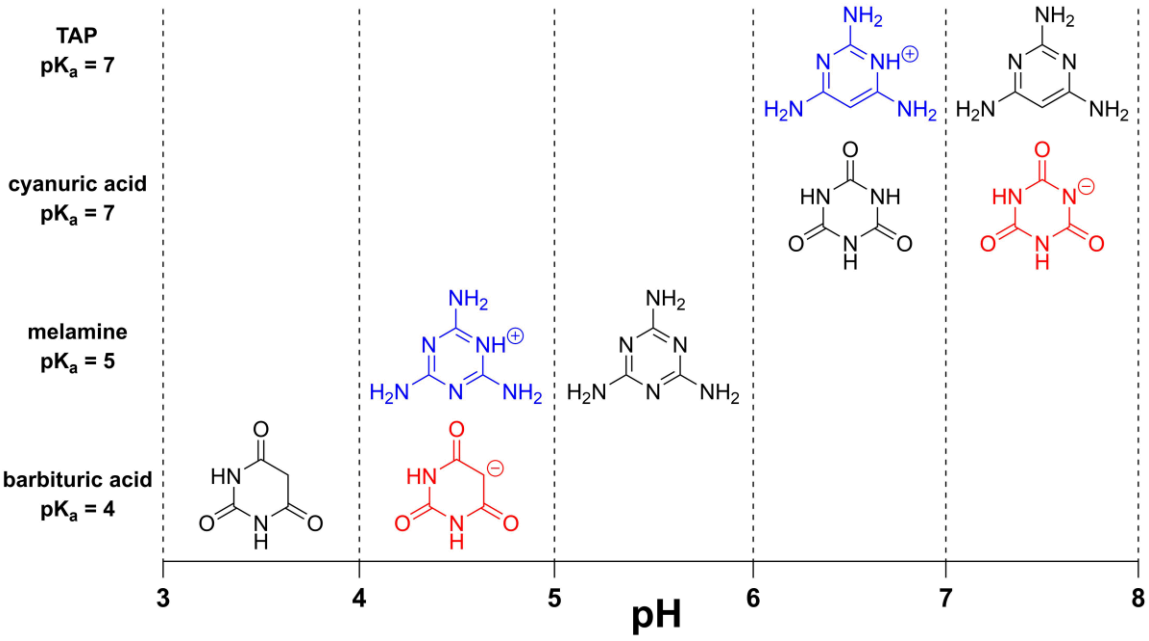


Figure 2.1. Charge states of the self-assembling noncanonical nucleobases according to pH. TAP and cyanuric acid, with the same pK_a value of 7, are expected to pair most strongly at pH 7 where a significant amount of each nucleobase is neutral. Melamine, with a pK_a of 5, is expected to pair well with cyanuric acid derivatives in the large range of pH 5-7, where both are found mostly in their neutral states. Barbituric acid, with a pK_a of 4, is expected to have weak pairing with melamine around pH 4.5, where some of each nucleobase is still neutral. Barbituric acid and TAP are not expected to pair, as there exists no pH where both are found in their neutral state in significant amounts.

Although the preliminary study of TARC (the β-*C*-ribofuranoside of TAP) demonstrated that supramolecular associations of a prebiotic noncanonical nucleoside with a complementary heterocycle at the monomer level were possible, it could not be expanded

to a two-nucleotide system in a prebiotically realistic scenario due to the insufficient reactivity of cyanuric acid. However, barbituric acid, the pyrimidine analogue of cyanuric acid, has been known to react with sugars in water to form glycosides. Utilizing glycosides of barbituric acid, however, introduces a new problem of pK_a incompatibility. The pK_a values of TAP and barbituric acid are 7 and 4, respectively. This means that there exists no pH value that TAP and barbituric acid would both exist in their neutral forms in an appreciable quantity; a condition thought to be required for hexad formation in water [6]. This problem can, to an extent, be alleviated by the substitution of TAP for melamine. In the pH vicinity of 4-5, melamine ($pK_a = 5$) and barbituric acid will both exist in their neutral states to some extent, allowing for base pairing. These pH considerations are summarized in Figure 2.1. Unlike TAP, melamine is not expected to form glycosides linked directly to the core heterocycle, but rather through its exocyclic amino groups (see § 2.3.1). The use of a barbituric acid-melamine system is also attractive for its prebiotic plausibility, as these heterocycles are formed from urea in the same model prebiotic reaction [7].

2.3.1 Prebiotic Syntheses of Nucleotides of Barbituric Acid, Melamine, and TAP

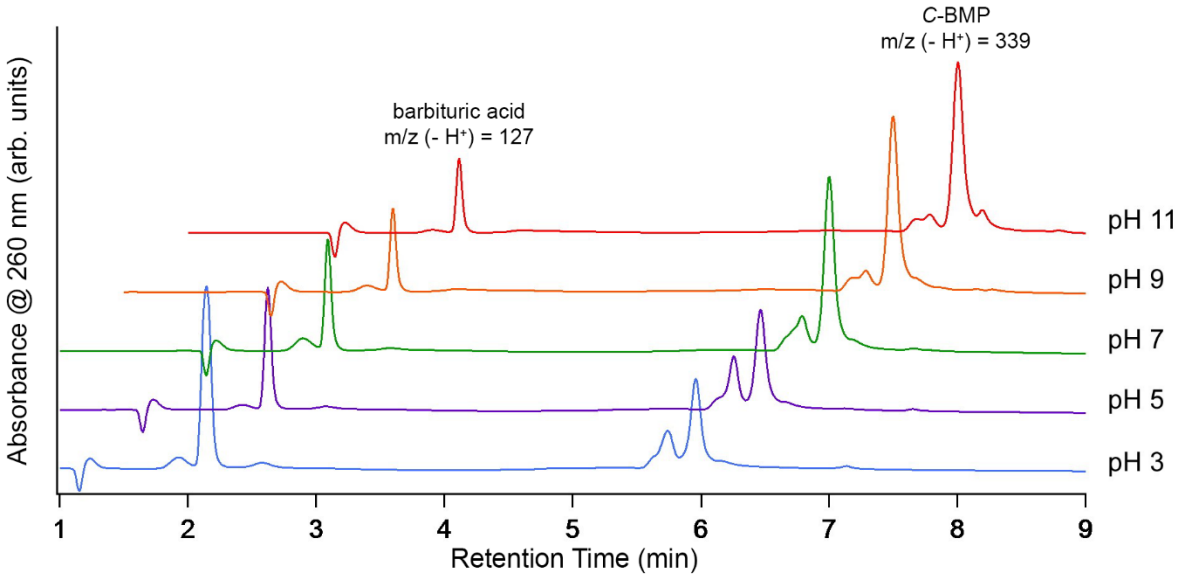


Figure 2.2. UV-LC/MS chromatograms of the reaction of barbituric acid (0.5 M) with ribose-5-phosphate (0.5 M) at room temperature for four hours at variable pH. An offset in the horizontal axis of 0.5 min was used for clarity.

The reaction of barbituric acid with sugars in water has been described [4, 8, 9]. Base is added in order to generate the active nucleophile, the barbiturate anion. The reaction of barbituric acid with ribose or ribose-5-phosphate in water was studied under different pH conditions to optimize yield (Figure 2.2). Unsurprisingly, the reaction proceeds to a greater yield at higher pH values. After reaching pH 10 (by the addition of sodium hydroxide), however, the solubility of barbituric acid begins to decrease, possibly due to the insolubility of the disodium dianion salt. Therefore, reactions of barbituric acid were performed at pH 9 (by the addition of sodium hydroxide). Room temperature was found to be sufficient for reactivity without leading to unintended side products. The maximum yield achieved in the synthesis of the nucleotide of barbituric acid (referred to as *C*-BMP,

Figure 2.4A) from barbituric acid and ribose-5-phosphate was 82%, with about 70% of the product formed as the β-anomer (determined by ROESY NMR spectroscopy) [10].

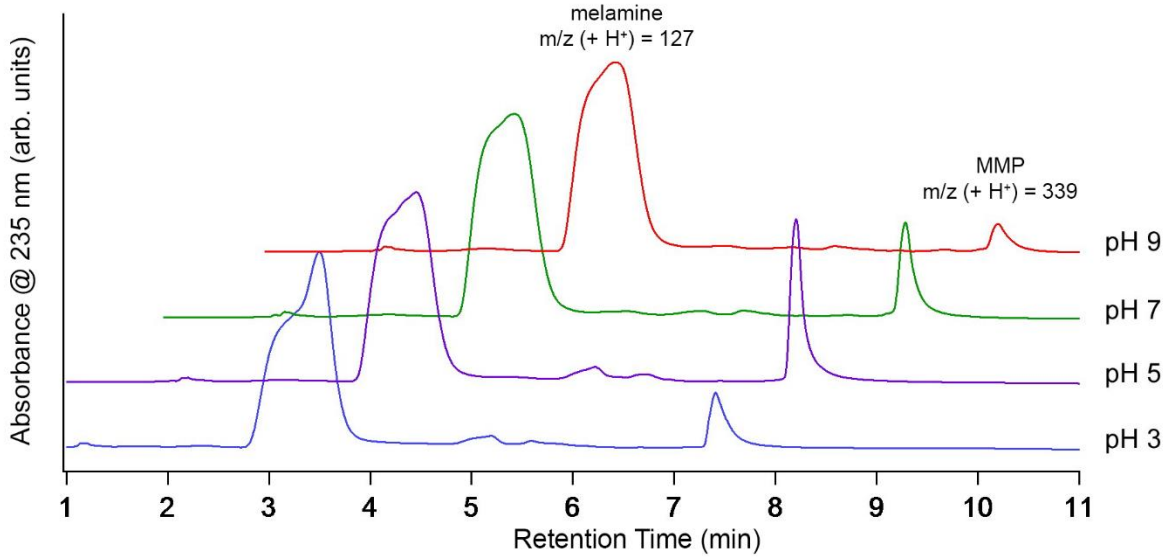


Figure 2.3. UV-LC/MS chromatograms of the reaction of melamine (0.2 M) with ribose-5-phosphate (0.2 M) at 65°C for 24 hours at variable pH. An offset in the horizontal axis of 0.5 min was used for clarity.

Melamine is expected to react with sugars via a Schiff base intermediate; a process which is known to be acid-catalyzed. A pH optimum of 5 (achieved by the addition of HCl) was found for the reaction of melamine with ribose (Figure 2.3); this pH value was also influenced by the solubility of melamine, which is limited at higher pH values. In order for the reaction to proceed quickly, a temperature of 65°C is required. The maximum yield achieved in the synthesis of the nucleotide of melamine (referred to as MMP, Figure 2.4B) from melamine and ribose-5-phosphate was 55% [10]; however, yields were typically in the vicinity of 35%. The reaction initially produces an excess of the α-anomer (60%), but this equilibrates to 45% after purification. This isomerization is probably due to a Curtin-Hammett effect. The reaction of melamine with ribose-5-phosphate can also proceed at

room temperature if allowed to react for several weeks; an inconvenient amount of time for laboratory purposes, but a relatively short amount of time for prebiotic chemistry.

Purification of *C*-BMP and MMP was achieved by ion exchange chromatography. MMP, with its basic melamine moiety, was purified by cation exchange chromatography. *C*-BMP, with several acidic sites, was purified by anion exchange chromatography; however, this purification was not robustly reproducible.

During the initial investigation of these noncanonical nucleotides, the glycosylation of TAP by ribose-5-phosphate to form the nucleotide, pTARC, was also investigated. Unlike the dry-down reaction that was used to form TARC, solution phase reactions were investigated for pTARC. In these reactions, a mixture of *N*-exocyclic and *C*-endocyclic nucleotides were formed. It was found that the *C*-nucleotide could be formed almost exclusively (obtained as both anomers) if the reaction is performed at pH 1 (Figure 2.4C). This was rationalized by considering the mechanisms of formation and reversion of the *C*- and *N*-nucleotides, and concluding that the reversion of the *C*-nucleotide to the sugar and base would proceed much more slowly than the reversion of the *N*-nucleotide. It was also found that the anomers could be separated by cation exchange chromatography.

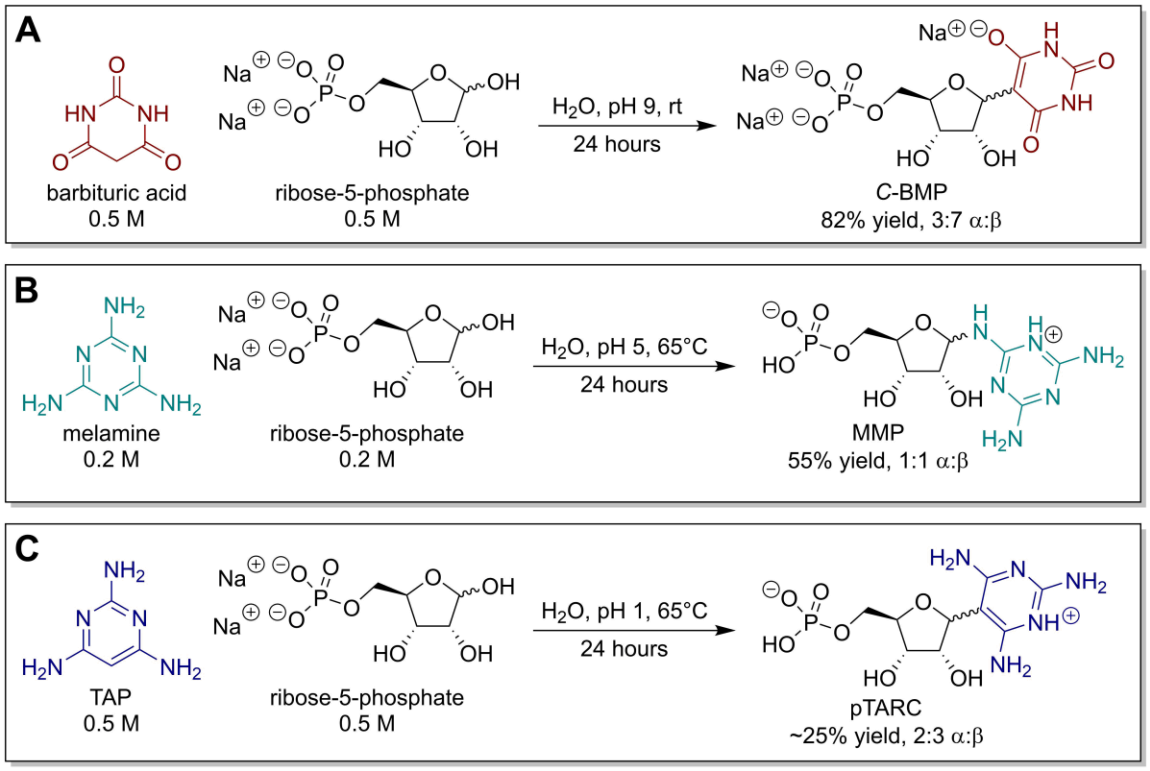


Figure 2.4. The optimized syntheses of the noncanonical nucleotides of A. barbituric acid, B. melamine, and C. TAP.

2.3.2 The Mechanism of Aqueous Glycosylation of Noncanonical Nucleobases

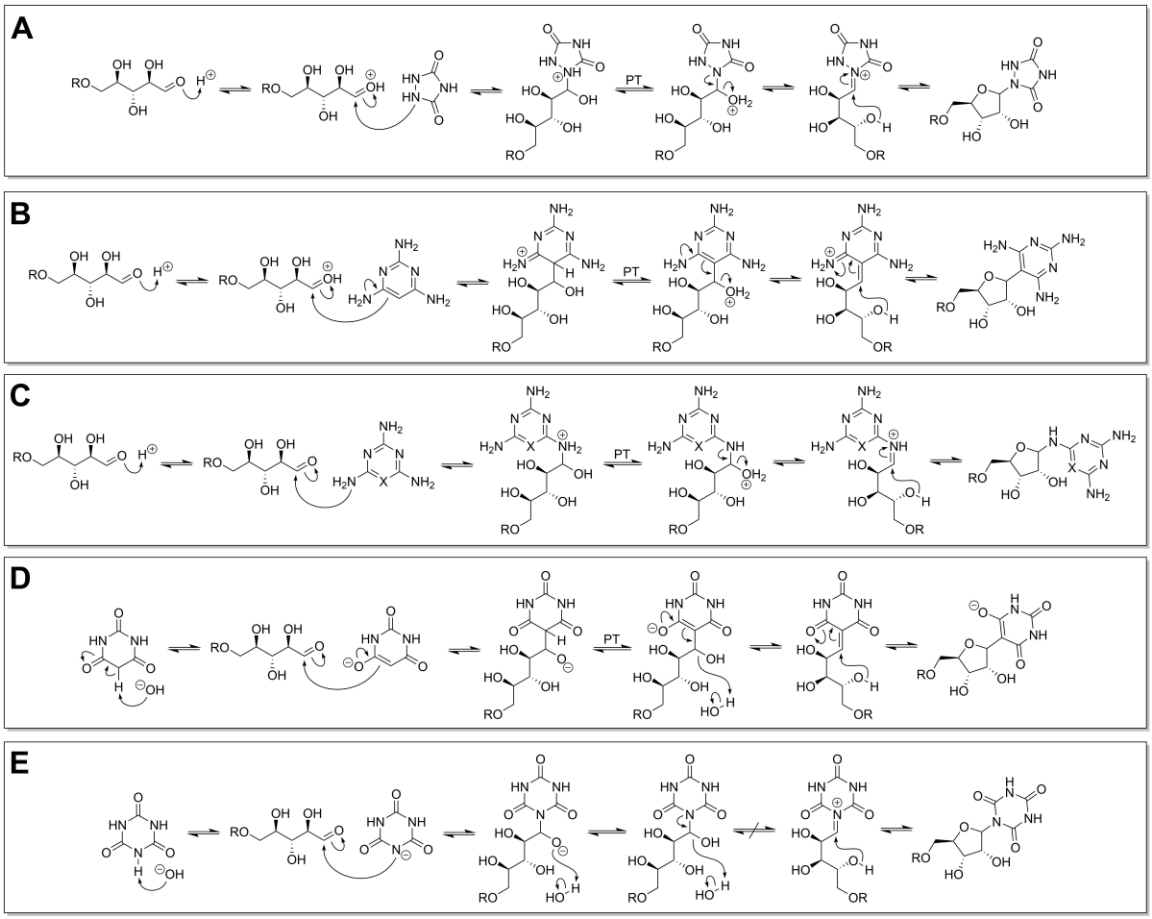


Figure 2.5. The mechanism of glycosylation of several noncanonical nucleobases with ribose (R = H) or ribose-5-phosphate (R = PO_3H_2). A. The acid-catalyzed glycosylation of urazole proposed by Dworkin and Miller. B. The acid-catalyzed *C*-glycosylation of TAP, first described by Chen et al. C. The acid-catalyzed *N*-glycosylation of TAP (X = CH) or melamine (X = N). D. The base-catalyzed glycosylation of barbituric acid. E. The failed glycosylation of cyanuric acid.

The ability of some noncanonical nucleobases to react with ribose in water, and the inability of the canonical nucleobases to do so, raises the question of the chemical mechanism of ribosylation. Dworkin and Miller offered a mechanism for the condensation of urazole with ribose that invokes the linear, free aldehydic form (Figure 2.5A) [11].

Assuming that attack of urazole on this aldehyde was the rate-limiting step of glycosylation, Dworkin and Miller showed that the rate of addition of urazole to various sugars at pH 4.5 correlates well with the amount of free aldehyde measured by different means, supporting the proposed mechanism.

A similar mechanism is probably operative for the addition of TAP (Figure 2.5B), melamine (Figure 2.5C), and barbituric acid (Figure 2.5D) to ribose and other sugars. For the basic noncanonical nucleobases TAP and melamine, glycosylation is probably acid-catalyzed and initiated by protonation of the free aldehyde of ribose. (This is supported, for example, by the optimum pH of 5 observed for the addition of melamine to ribose-5-phosphate.) Acting as a *C*-nucleophile, the glycosylation of TAP is reminiscent of electrophilic aromatic substitution. Attack by C5 of TAP generates an intermediate that temporarily breaks aromaticity, but is stabilized by the three electron-donating amino substituents. After a solvent-mediated proton transfer, aromaticity is restored, and then temporarily broken again to eliminate water, forming a Knoevenagel-like intermediate. Cyclization by internal nucleophilic attack by one of the hydroxyl groups of ribose gives the nucleoside product.

The mechanism is similar for condensation with an exocyclic amino group (Figure 2.5C) of TAP (X = CH) or melamine (X = N). In an acid-catalyzed manner, an iminium intermediate is generated with the release of water, which then cyclizes to give the nucleoside.

Glycosylation of barbituric acid, unlike the previous examples, probably occurs in a base-catalyzed manner (Figure 2.5D), as the reaction tends to proceed more efficiently at

higher pH values. Deprotonation of barbituric acid gives the nucleophilic barbiturate anion, which attacks the free aldehyde of ribose and proceeds to a Knoevenagel intermediate in a manner reminiscent of TAP *C*-glycosylation.

In all examples shown so far, the formation of a double-bonded intermediate, formed by the condensation of the nucleobase with the free aldehyde of ribose, is crucial to nucleoside formation. It is for this reason that cyanuric acid (and the canonical nucleobases) fails to react with ribose in water (Figure 2.5E). The hypothetical double-bonded intermediate for cyanuric acid would be highly destabilized, as the lone pair of the glycosidic nitrogen atom is withdrawn by the adjacent oxo groups.

The importance of this glycosylation mechanism, and other mechanisms of appending nucleobases to candidate proto-nucleic acid backbones, will be discussed in further detail in Chapter 4.

2.3.3 Supramolecular Assembly of Noncanonical Nucleotides

The supramolecular assembly of heterocycles to form hexameric rosettes in water is often accompanied by hydrogelation [6, 10, 12]. Therefore, the first assessment of whether or not these nucleotides could assembly with a pairing partner was attempts at hydrogel formation. It was found that purified MMP, when incubated with barbituric acid at 50 mM in each monomer, pH 4.5, 300 mM NaCl, at 4°C overnight, forms a hydrogel. *C*-BMP, whether crude or purified, easily forms gels with melamine at pH 4.5, but also, unexpectedly, at pH 7. As the pK_a of the parent heterocycle barbituric acid is 4, and assuming that the pK_a does not significantly change by glycosylation, the recognition unit of *C*-BMP is expected to be negatively charged, which is thought not to be able to assemble.

When purified C-BMP and MMP are combined, no hydrogel formation is observed at any pH.

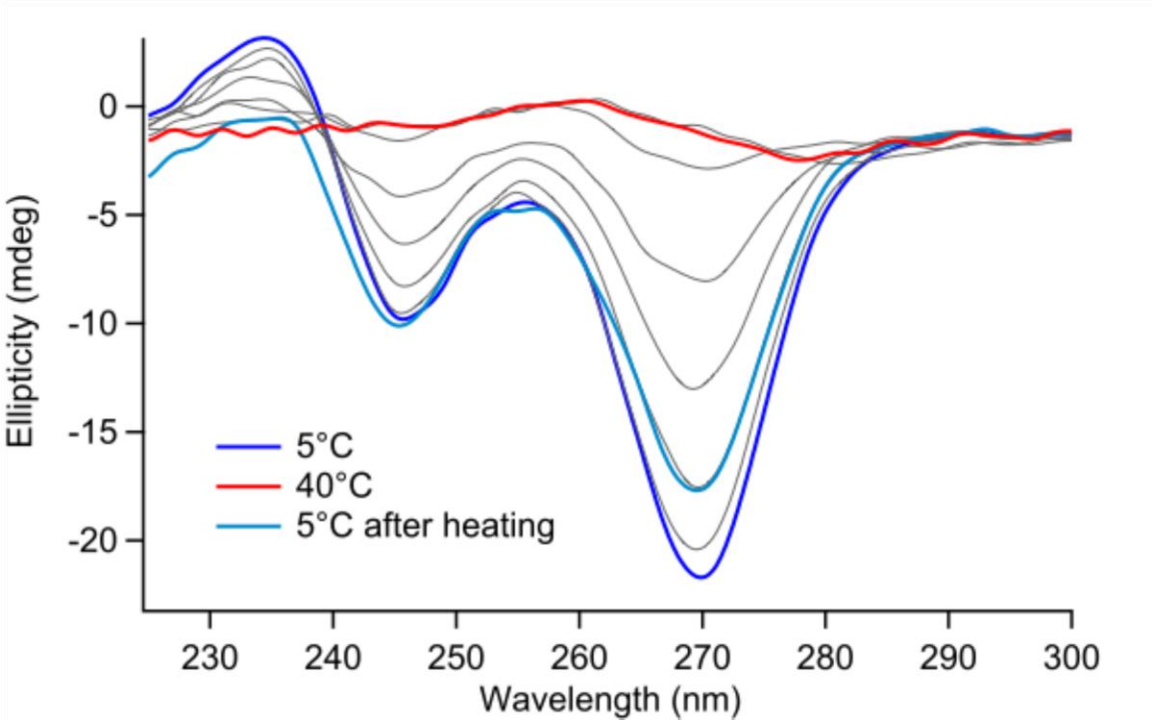


Figure 2.6. CD analysis of a solution containing *C*-BMP and melamine. CD spectra of a solution (50 mM in each molecule; 1 M NaCl; pH 4.5), ranging in temperatures between from 5 to 40°C, and at 5°C after being heated twice to 40°C. Loss of signal when heated to 40°C and return of signal when cooled to 5°C illustrates the non-covalent nature of the assemblies formed by *C*-BMP and melamine. Note that change in intensity with heat cycling is due to the kinetic behavior of supramolecular assembly nucleation and growth, which results in variations in the amount of chiral assembly formed for the same sample when cooled to 5°C after heating.

These early hydrogelation assays were supplemented with circular dichroism (CD) measurements to determine whether or not supramolecular assembly was occurring without any associated hydrogelation. CD spectra were collected for the MMP/*C*-BMP system, the MMP/barbituric acid system, and the *C*-BMP/melamine system (Figure 2.6). While all systems gave enhanced signals with respect to their isolated chiral monomers, the shape of the spectra were not reproducible in general.

2.3.4 Attempts at Polymerization of the Noncanonical Nucleotides by Chemical Activation

When the noncanonical nucleotides assemble in water, the fibers formed are often micrometer length (as detected by atomic force microscopy), corresponding to tens of thousands of ordered monomers [10]. This assembly could be used to polymerize the monomers with condensing agents. While the prebiotic plausibility condensing agents is dubious (the ones tested, such as 1-cyanoimidazole, are not prebiotically plausible, but others, such as cyanoacetylene [13] and dicyandiamide [14], may have been transiently present on the early earth), the demonstration of the ability to form polymers from a noncovalent assembly is nevertheless a milestone in origin-of-life research.

Polymerization was first attempted by cyanogen bromide on the MMP/barbituric acid and *C*-BMP/melamine systems. Chemical activation by cyanogen bromide is expected to proceed by nucleophilic attack of phosphate, followed by elimination of bromide to form a phosphate-cyanate mixed anhydride. This activated phosphate is then attacked by a hydroxyl group of an adjacent nucleotide (at the 2′ or 3′ position) with concerted loss of cyanate to form a phosphodiester. The reactions were performed by forming hydrogels (pH 5) and adding a freshly prepared concentrated solution of cyanogen bromide. Controls were performed with the unassembled nucleotides in solution without a pairing partner. A DNA ligation positive control with 12-mer tiling system was also performed at pH 7.5 and pH 5. All reactions were performed in PCR tubes at 4°C overnight. DNA ligation was observed only in the pH 7.5 sample, and no oligomerization was found in either hydrogel sample. However, brominated barbituric acid was detected by mass spectrometry in both hydrogel reactions, suggesting a side reaction that produced cyanide. The gel property of the solution

was also lost during the course of the reaction, and the pH was found to have dropped to 2.5, probably due to hydrolysis of cyanogen bromide to form hydrobromic acid. A similar set of reactions was prepared using the condensing agent 1-cyanoimidazole; however, no ligation or oligomerization was observed.

2.4 Barbituric Acid Glycosides

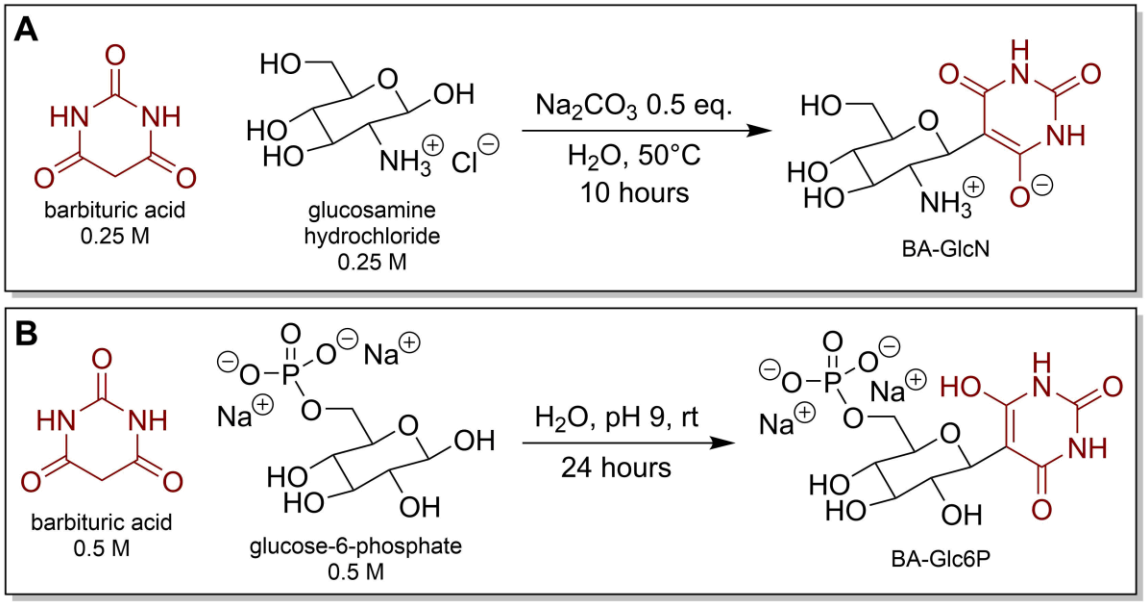


Figure 2.7. Examples of glycosylation reactions of barbituric acid. A. The reaction of barbituric acid with glucosamine produces the β-*C*-glucopyranoside, BA-GlcN. B. The reaction of barbituric acid with glucose-6-phosphate produces the β-*C*-glucopyranoside, BA-Glc6P.

The reaction of sugars with barbituric acid in water is facile and proceeds to high yields. Barbituric acid is a relatively strong carbon acid (pK_a = 4) due to the aromatic stabilization of the barbiturate anion. Barbiturate is a carbon nucleophile, readily participating in Knoevenagel reactions in water.

One particularly interesting reaction is the glycosylation of barbituric acid with glucosamine (Figure 2.7A). Like other glycosylation reactions of barbituric acid, this

reaction is base-catalyzed. This reaction has been described in the literature [4]. The main product is the *C*-β-glucopyranoside (denoted BA-GlcN), which is expected to be zwitterionic at neutral pH. Consistent with this, it has a low solubility in water; a common property for zwitterionic compounds. In fact, it is purified by recrystallization in water.

Unlike barbituric acid, TAP and melamine fail to react with glucosamine under near-neutral pH conditions. Instead, the reaction rapidly browns, presumably by the reaction of glucosamine with itself: two molecules of glucosamine condense to form a dihydropyrazine moiety, which can possibly be oxidized by air to give pyrazine products that may undergo further elimination reactions to produce conjugated, insoluble tars with high extinction coefficients [15]. Presumably, of the noncanonical nucleobases studied, only barbituric acid is a strong enough nucleophile to capture glucosamine before it decomposes.

The compound BA-GlcN, when dissolved with melamine in water at 50 mM each, at pH 4.5, forms a hydrogel, indicative of fibrous stacked-hexad supramolecular assembly. To date, BA-GlcN is the only compound that successfully forms soluble hexad assemblies with a peripheral positive charge (endowed by the protonated amino group of glucosamine); all other molecules that successfully assemble (with an underivatized complementary heterocycle) have a peripheral negative charge.

Barbituric acid also reacts with other glucose derivatives to give β-glucopyranosides, a number of which have also been previously described [4, 8, 9]. This selectivity for sugar conformation, while not absent in certain other sugars, is especially pronounced for glucose and its derivatives, and is exemplified well by the reaction of

barbituric acid with different sugars. For example, the gluconucleotide of barbituric acid (formed from the reaction of glucose-6-phosphate with barbituric acid) is produced exclusively as the β-pyranoside (Figure 2.7B). However, the ribonucleotide (*C*-BMP) is produced as a mixture of the α- and β-furanoside anomers (Figure 2.4A). (Pyranoside forms are not possible for ribose-5-phosphate.) This selectivity in glucose is attributable to the stereochemical configuration of the substituents of glucopyranose, which all occupy equatorial positions (except for the allowed variability in at the anomeric position). The substitution of the anomeric position with a sterically large group, such as barbituric acid, frustrates the formation of the α-pyranoside: analogous to the substitution of cyclohexane with a tert-butyl group, the steric bulk of an endocyclic *C*-nucleobase substituent prevents it from occupying an axial position. Therefore, in order for a hypothetical α-glucopyranoside of barbituric acid to exist, a chair flip must occur to bring the barbituric acid moiety to an equatorial position; however, this places all other glucose substituents in axial positions, which is also sterically disfavored. This stability of β-glucopyranosides also has important consequences for glycosidic bond cleavage. In practice, it is observed that purification of *C*-BMP is complicated by nucleotide degradation by glycosidic bond cleavage. However, the analogous gluconucleotide of barbituric acid does not suffer from the problem.

The ability of barbituric acid glycosides to assemble with melamine is somewhat paradoxical with respect to empirically determined rules for stacked-hexad supramolecular assembly. The two rules in question are that 1) the assembling heterocycle must not be charged, and 2) the assembling heterocycle must be planar. For the most prevalent tautomer of barbituric acid, the keto tautomer, meeting these two criteria simultaneously seems

impossible, because the C5 position of barbituric acid is tetrahedral when barbituric acid is protonated and neutral. It may be, however, that glycosylation ameliorates this problem by presenting the endocyclic oxygen atom of the sugar as a hydrogen bond acceptor that stabilizes the planar enol form of a neutral barbituric acid moiety. The requirement of the enol tautomer of barbituric acid for supramolecular assembly is also supported by the observation that the formation of supramolecular assemblies by underivatized barbituric acid and MMP (the nucleotide of melamine) is slow. Unlike other systems studied, which form hydrogels within several seconds when both assembling components are present in solution at the appropriate concentration and pH, hydrogels of barbituric acid and MMP only occur after a solution of the two has been incubated at 4°C overnight. Self-assembly of barbituric acid may shift the tautomeric equilibrium in favor of the enol, but reaching this new equilibrium has an appreciable activation barrier.

2.5 Melamine Glycosides

Of the hexad-forming noncanonical nucleobases that successfully form glycosides, melamine is the least reactive (i.e., its reactions with sugars tend to proceed only to low yields). This is not surprising, as it is less electron-rich than TAP or the barbiturate anion. While isoelectronic with TAP, the replacement of the CH moiety of the pyrimidine to the N moiety of the triazine makes the heterocyclic ring more electron-withdrawing; therefore, the exocyclic amino groups are less nucleophilic. Additionally, melamine is poorly soluble in water at pH values above 5 (the pK_a of melamine). Even at pH 5, melamine can be dissolved in water to only about 0.2 M at 65°C. The poor nucleophilicity and low solubility of melamine both contribute to the typically low yields (< 50%) of glycosylation reactions.

The reaction of melamine with certain sugar derivatives can also be complicated by co-precipitation. Melamine, when protonated, tends to co-precipitate with oxyanions; e.g., phosphates, sulfates, and carboxylates. For example, the reaction of melamine with glucuronic acid in water (0.2 M each, with no pH adjustment) is complicated by the co-precipitation of melamine with glucuronic acid. Interestingly, the reaction of melamine with galacturonic acid (the C-4 epimer of glucuronic acid) proceeds very differently. The solubility of melamine in solutions containing galacturonic acid is anomalously high. When present in equimolar amounts, melamine can be fully dissolved at 1 M with heating. Upon cooling, this concentrated solution forms a thick, clear hydrogel with a pale pink color.

2.6 Introduction to TAP Glycosides

The reaction of 2,4,6-triaminopyrimidine (TAP) with sugars is more complicated than the reactions of melamine or barbituric acid due to the presence of three distinguishable nucleophilic sites on TAP: the exocyclic amino group at position 2, the exocyclic amino group at position 4/6 (equivalent by symmetry), and the endocyclic carbon atom at position 5 (see Figure 1.4). The amino groups at positions 4 and 6 are expected to be more nucleophilic than the amino group at position 2, as the amino group at position 2 is *ortho* to two endocyclic, electron-withdrawing nitrogen atoms, while the amino groups at positions 4 and 6 are *ortho* to only one endocyclic nitrogen atom. Substitution of TAP at position 5 (by electrophilic aromatic substitution) is predicted to be kinetically slow. This can be inferred by the rate of proton exchange of the hydrogen atom at position 5 when TAP is dissolved in deuterated water (D_2O): after TAP is dissolved, all of the amino group protons are rapidly exchanged, and cannot be seen in the ^1H NMR spectrum.

However, the proton at position 5 exchanges slowly enough at neutral and basic pH values that a peak can be detected in the ^{1}H NMR spectrum [16].

The site of glycosylation of TAP is dependent on the mode of the reaction (solution phase or dry-down), the pH of the reaction, and the identity of the sugar. For example, when TAP is reacted with ribose by drying from an aqueous solution at 35°C with an initial pH of approximately 8, a mixture of *N*- and *C*-ribosides are formed, with TARC (the β-*C*-ribofuranoside of TAP) being produced in the highest yield [5]. Similarly, if TAP is reacted with ribose-5-phosphate in water at 0.5 M in each reactant, pH 7, at 65°C for 24 hours, a mixture of *N*- and *C*-nucleotides are formed. However, if this reaction is performed at pH 1, the α- and β-*C*-nucleotides (i.e., α- and β-pTARC) are produced almost exclusively (see § 2.3.1 and Figure 2.4).

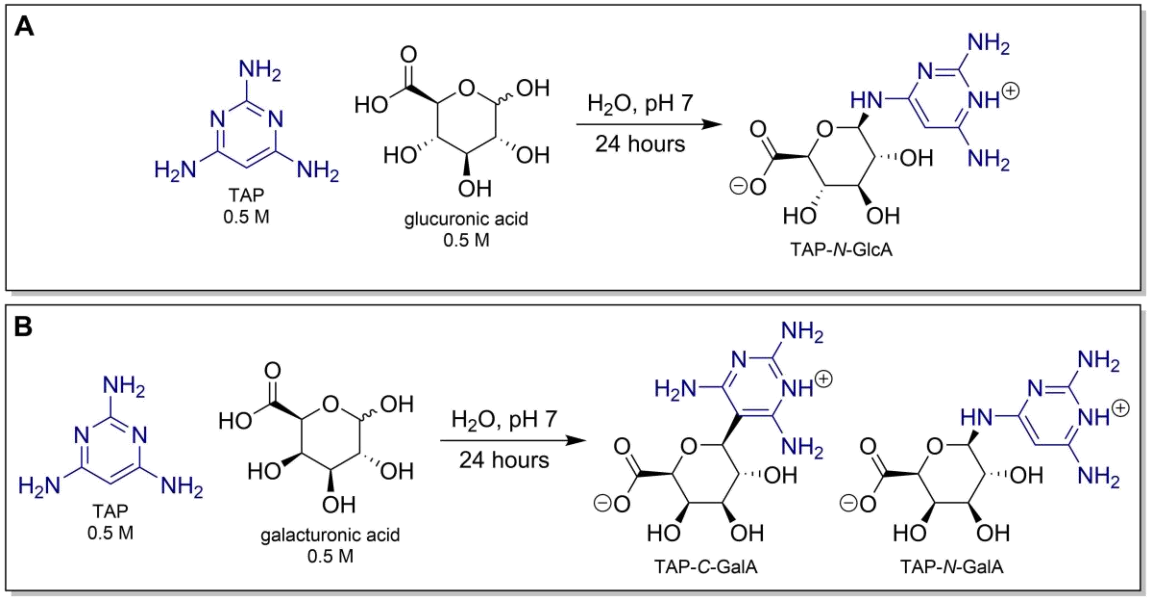


Figure 2.8. Glycosylation reactions of TAP with uronic acids. A. The reaction of TAP with glucuronic acid produces the β-4-*N*-glucopyranoside, TAP-*N*-GlcA. B. The reaction of TAP with galacturonic acid produces both the β-*C*-glucopyranoside, TAP-*C*-GalA, and the β-4-*N*-glucopyranoside, TAP-*N*-GalA.

Like the reactions of melamine with uronic acids, the reactions of TAP with uronic acids is complicated by the formation of insoluble co-precipitates. When equimolar amounts of solid TAP and glucuronic acid or galacturonic acid are suspended in water to produce a 0.5 M solution (in each component), the two components will briefly form a solution (with heating and vortexing), but then rapidly form a white, amorphous co-precipitate. Further heating at approximately 95°C (on a heating block) will eventually bring the two components into solution, and after adjusting the pH to 7, leaving this solution at 65°C for 24 hours will produce a new precipitate which is the zwitterionic uronoside(s) of TAP. For the reaction of TAP with glucuronic acid, only the β-4-*N*-glucuronoside (i.e., the glucuronopyranoside of TAP where TAP is substituted at the exocyclic amino group at position 4, denoted TAP-*N*-GlcA) was found in this precipitate (Figure 2.8A). However, the reaction of TAP with galacturonic acid, under the same conditions, produced both the β-4-*N*-galacturonoside (denoted TAP-*N*-GalA) and the β-5-*C*-galacturonoside (denoted TAP-*C*-GalA) in a 1:1 ratio in the precipitate (Figure 2.8B). Interestingly, when pure TAP-*N*-GlcA is dissolved in water at 50 mM at pH 7 with one equivalent of cyanuric acid, gelation does occur. Instead, a precipitate is formed, presumably from both components, as TAP-*N*-GlcA is expected to be anionic (rather than zwitterionic) at this pH. However, the mixture of TAP-*N*-GalA and TAP-*C*-GalA, when dissolved in water at 50 mM (approximately 25 mM in each glycoside) at pH 7 with one equivalent (i.e., 50 mM) of cyanuric acid, a hydrogel is formed.

These eccentricities in the reactivity of TAP with different sugars and in the propensity of these glycosides to form soluble assemblies may provide criteria for the selection of some pre-RNA monomers over others.

2.7 References

[1] Fuller, W.D., R.A. Sanchez, and L.E. Orgel, *Studies in prebiotic synthesis. VI. synthesis of purine nucleosides.* J. Mol. Biol., 1972. **67**(1): p. 25-33.

[2] Fuller, W.D., R.A. Sanchez, and L.E. Orgel, *Studies in prebiotic synthesis: VII. solid-state synthesis of purine nucleosides.* J. Mol. Evol., 1972. **1**: p. 249-57.

[3] Kolb, V.M., J.P. Dworkin, and S.L. Miller, *Alternative bases in the RNA world: The prebiotic synthesis of urazole and its ribosides.* J. Mol. Evol., 1994. **38**: p. 549-557.

[4] Gonzalez, M.A., et al., *Facile preparation of C-glycosylbarbiturates and C-glycosylbarbituric acids.* Carbohydrate Research, 1986. **158**: p. 53-66.

[5] Chen, M.C., et al., *Spontaneous prebiotic formation of a β-ribofuranoside that self-assembles with a complementary heterocycle.* Journal of the American Chemical Society, 2014. **136**(15): p. 5640-5646.

[6] Cafferty, B.J., et al., *Ultra-sensitive pH control of supramolecular polymers and hydrogels: pK(a) matching of biomimetic monomers.* Chemical Science, 2014. **5**(12): p. 4681-4686.

[7] Menor-Salvan, C., et al., *Synthesis of pyrimidines and triazines in ice: Implications for the prebiotic chemistry of nucleobases.* Chemistry-a European Journal, 2009. **15**(17): p. 4411-4418.

[8] Martinez, M.B., et al., *Synthesis and conformational-analysis of C-glycosylbarbiturates.* Carbohydrate Research, 1990. **199**(2): p. 235-238.

[9] Wulff, G. and G. Clarkson, *On the synthesis of C-glycosyl compounds containing double-bonds without the use of protecting groups.* Carbohydrate Research, 1994. **257**(1): p. 81-95.

[10] Cafferty, B.J., et al., *Spontaneous formation and base pairing of plausible prebiotic nucleotides in water.* Nature Communications, 2016. **7**: p. 11328.

[11] Dworkin, J.P. and S.L. Miller, *A kinetic estimate of the free aldehyde content of aldoses.* Carbohydr. Res., 2000. **39**: p. 359-365.

[12] Cafferty, B.J., et al., *Efficient self-assembly in water of long noncovalent polymers by nucleobase analogues.* Journal of the American Chemical Society, 2013. **135**(7): p. 2447-2450.

[13] Ferris, J.P., *Cyanovinyl phosphate: a prebiological phosphorylating agent?* Science, 1968. **161**(3836): p. 53-4.

[14] Hulshof, J. and C. Ponnamperuma, *Prebiotic condensation-reactions in an aqueous-medium - review of condensing agents.* Origins of Life and Evolution of the Biosphere, 1976. **7**(3): p. 197-224.

[15] Hrynets, Y., M. Ndagijimana, and M. Betti, *Studies on the Formation of Maillard and Caramelization Products from Glucosamine Incubated at 37 °C.* Journal of Agricultural and Food Chemistry, 2015. **63**(27): p. 6249-6261.

[16] Fialho, D.M., et al., *Glycosylation of a model proto-RNA nucleobase with non-ribose sugars: implications for the prebiotic synthesis of nucleosides.* Organic & Biomolecular Chemistry, 2018. **16**: p. 1263-1271.

CHAPTER 3. GLYCOSYLATION OF 2,4,6-TRIAMINOPYRIMIDINE (TAP) WITH NON-RIBOSE SUGARS

3.1 Introduction

Unlike the canonical nucleobases of RNA, 2,4,6-triaminopyrimidine (TAP) has the ability to react with ribose [1], and other sugars [2], in water to form glycosides. Furthermore, TAP and its derivatives have the ability to undergo supramolecular assembly in water with a complementary heterocycle, such as cyanuric acid, to form micron-length stacked-hexad fibers that resemble nucleic acids in their morphology [1, 3]. For this reason, TAP is considered a strong candidate pre-RNA nucleobase.

3.2 Experimental Procedures

3.2.1 Materials

2,4,6-Triaminopyrimidine (TAP) and D-mannose were purchased from Acros Organic. D-Galactose was purchased from Alfa Aesar. L-Ribulose was purchased from ZuChem. All other chemicals were purchased from Sigma Aldrich. All chemicals were used as received.

3.2.2 Reactions of TAP with Sugars

In a 1.5 mL microcentrifuge tube, 125 mg TAP (1 mmol) and 1 mmol of sugar were added. The solution was first dissolved in the appropriate amount of 5 M HCl to bring the solution to pH 7 or pH 1, and then deionized water was added to bring the total volume to 1 mL. The tubes were then sealed and heated in an oven at 85°C for 24 hours.

3.2.3 Analytical Techniques

NMR analysis: The crude reaction products of TAP with each sugar was first centrifuged to pellet any insoluble precipitates formed during the reaction. 60 μL were then taken from the supernatant and lyophilized. The samples were then redissolved in D_2O and lyophilized again. The samples were then dissolved in 600 μL of a sodium phosphate buffer in D_2O, 200 mM, pD 12, containing TSP 1 mM as an internal standard, and analyzed on a Bruker 800 MHz NMR spectrometer.

UV-LC/MS: A completed crude reaction of TAP with a sugar was first centrifuged to pellet any insoluble precipitates formed during the reaction. A sample was taken and diluted 1:1000 for LC/MS analysis. 2 μL samples were injected onto a 3.5 μm XBridge amide column running a linear gradient of 90% MeCN/10% NH_4OAc buffer 10 mM pH 9 to 60% MeCN/40% NH_4OAc buffer 10 mM pH 9 over a period of 7 minutes with a flow rate of 0.5 mL/min.

3.2.4 Purification of TAP Glucosides

The crude reaction mixtures of *N*-acetylglucosamine and glucose were purified on a Teledyne Isco CombiFlash Rf+ flash chromatography system using a Teledyne Isco C18Aq column with a 100% H_2O eluent and a 100% MeCN column wash after product elution. Fractions containing only the product masses (as determined by ESI-MS) were lyophilized and combined. The crude reaction mixture of TAP with glucose-6-phosphate was purified using SP Sephadex C-25 cation exchange resin with an isocratic NH_4OAc elution buffer, 50 mM, pH 4.3. Fractions containing only the product masses (as determined by ESI-MS) were lyophilized and combined.

3.3 Possible Prebiotic Synthesis of TAP

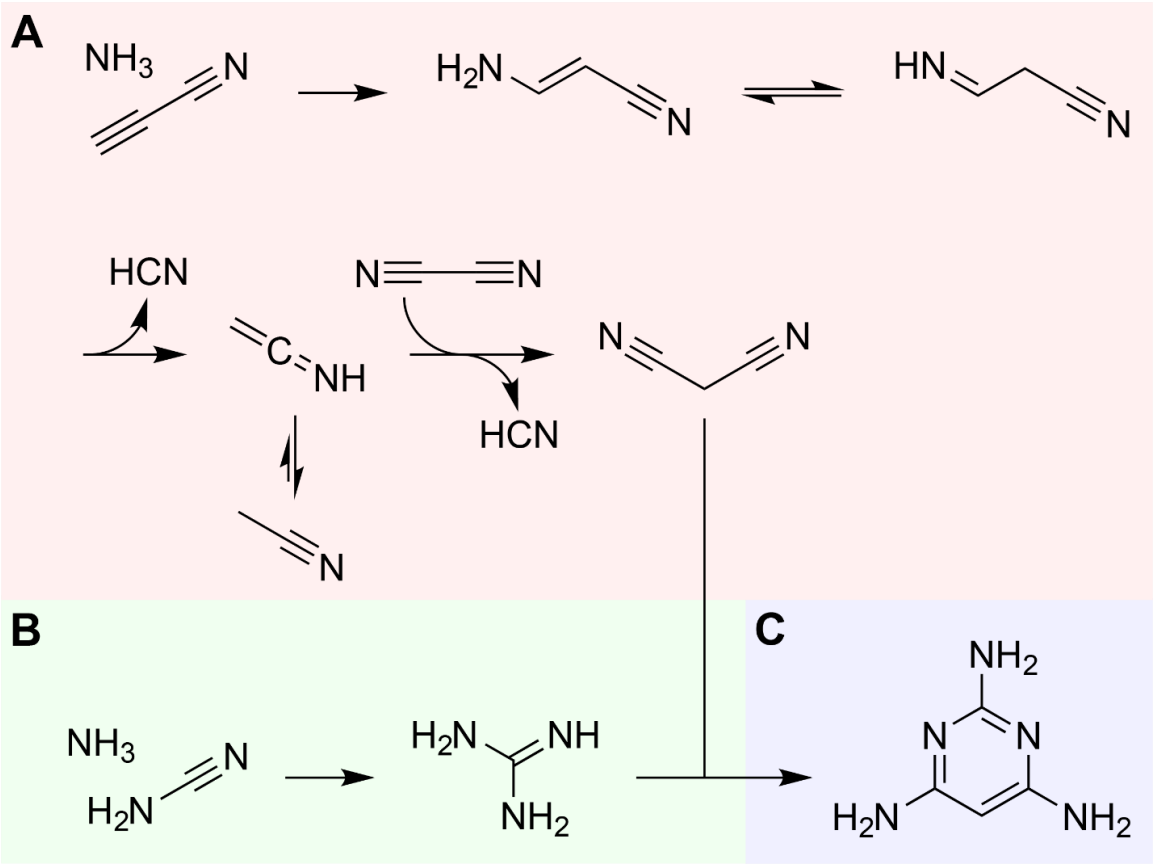


Figure 3.1. A possible prebiotic synthesis of TAP, originally suggested by Trinks and Eschenmoser [8]. A. Formation of malononitrile from ammonia, cyanoacetylene, and cyanogen. B. Formation of guanidine from ammonia and cyanamide. C. Formation of TAP from malononitrile and guanidine.

Neither a prebiotic synthesis nor detection in carbonaceous meteorites has been reported for TAP. TAP and its derivatives can be synthesized by conventional means by the reaction of guanidine with malononitrile (or derivatives thereof) [4, 5]. In particular, C5 derivatives of TAP can be synthesized by exploiting the nucleophilicity of the conjugate base of malononitrile. The α-position of malononitrile is a weak acid with a pK_a of 11 in water. Upon deprotonation, malononitrile can be monosubstituted or disubstituted at this

position. Monosubstituted derivatives of malononitrile can then be fused with guanidine to give 5-subtituted TAP compounds.

Guanidine and malononitrile have been invoked in prebiotic syntheses of substituted pyrimidines by Carell and coworkers [6, 7]. A hypothetical prebiotic synthesis of malononitrile has been suggested by Trinks and Eschenmoser [8] in which cyanoacetylene reacts with ammonia to form 3-aminoacrylonitrile, which tautomerizes and eliminates acetonitrile with the concomitant formation of hydrogen cyanide. The reaction of acetonitrile with cyanogen gives malononitrile with the release of hydrogen cyanide (Figure 3.1A). Details of the mechanism of this reaction are not given, but would presumably require an acid or base catalyst, or high temperatures, as might be found in atmospheric processes under UV radiation. Guanidine could be produced prebiotically from the reaction of ammonia with cyanamide (Figure 3.1B). The fusion of guanidine with malononitrile yields TAP (Figure 3.1C).

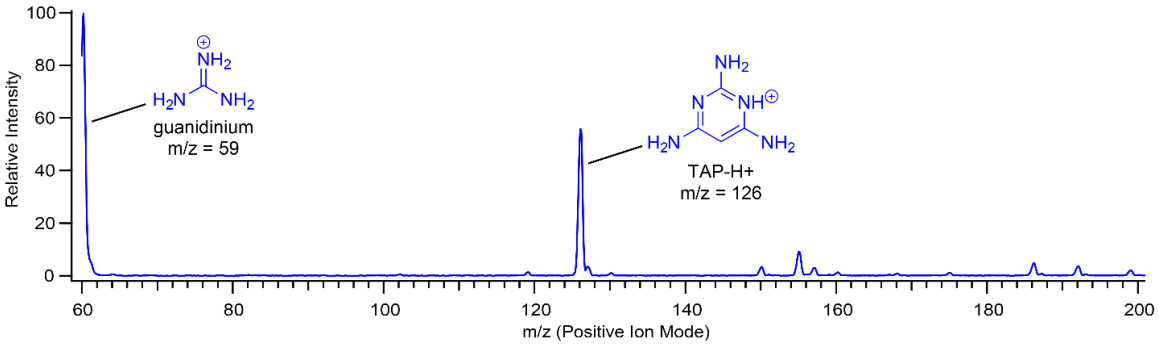


Figure 3.2. Mass spectrum, positive ion mode, of crude reaction mixture of malononitrile and guanidinium chloride 2:1, neat, 145°C, 15 hours. The 2,4,6-triaminopyrimidine product is seen in relatively high abundance at m/z = 126.

When a 1:2 solid mixture of guanidinium chloride and malononitrile are melted at 145°C for 15 hours, TAP is formed and detected by mass spectrometry (Figure 3.2). Although the compounds in this experiment were pure and concentrated, an analogous

nitrogen-rich prebiotic environment might also be able to produce TAP and related heterocycles. As mentioned previously, the selection for TAP over other nucleobases in pre-RNA may have occurred due to its enhanced reactivity towards sugars and the propensity of its derivatives for supramolecular self-assembly.

3.4 The Reactions of TAP with a Suite of Sugars

The lack of selectivity observed for prebiotic sugar-forming reactions, such as the formose reaction [9], and the ability of melamine (§ 2.5), barbituric acid (§ 2.4), and TAP (§ 2.6, and see below) to react with sugars other than ribose, suggests that the selection for ribose in RNA was not the result of a prebiotic predisposition based on chemical reactivity. To further reinforce this important principle, a study was conducted in which TAP was reacted with seventeen different sugars in water. In addition to the relevance of this study to prebiotic chemistry, we were interested to learn if the glycoside structural selectivity observed in reactions of TAP with ribose would also be exhibited by other sugars. These observations include chemoselectivity in the substitution of TAP (i.e., C5 versus exocyclic amine substitution), selectivity in the formation of either furanosides or pyranosides, and stereoselectivity in the orientation of TAP at the anomeric position of the sugar (either the α- or β-configuration). Structural characterization of glycosides formed by the reaction of TAP with glucose, glucose-6-phosphate, and *N*-acetylglucosamine, confirm that glycoside formation by TAP is not limited to ribose or ribose derivatives. Evidence of TAP glycosylation by several other hexose and pentose sugars, based on UV-LC/MS and ^{1}H NMR analyses, indicate that TAP reacts with a wide range of sugars to form aldosides. These observations have implications regarding the variety of glycosides (including nucleosides and proto-nucleosides) potentially present on the prebiotic Earth.

3.4.1 Selection of Sugars and Reaction Conditions

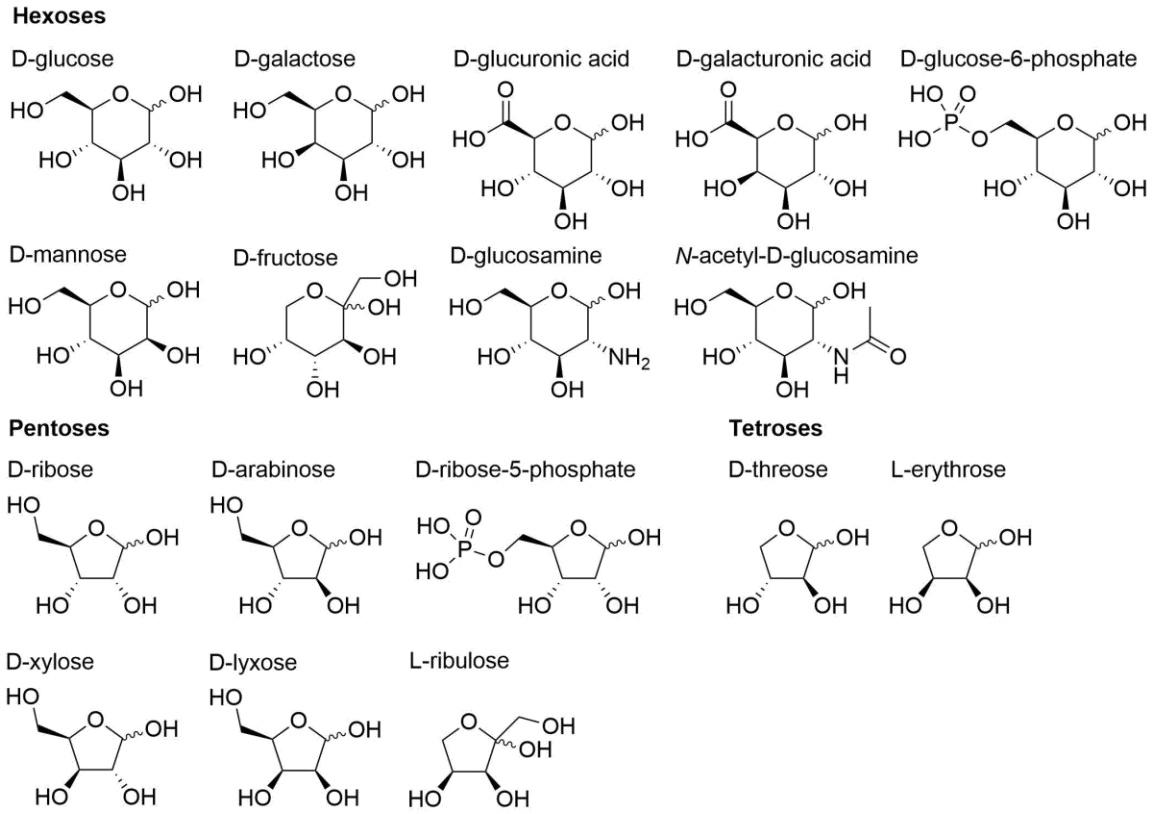


Figure 3.3. Sugars investigated in the current study for their reactivity with TAP. Note that hexoses are shown in their pyranose forms and pentoses in their furanose forms, but most sugars in solution exist in equilibrium between both ring forms and their open-chain aldehyde/ketone forms. Adapted with permission from [2].

Five aldopentoses, one ketopentose, eight aldohexoses, one ketohexose, and two aldotetroses were investigated for their potential to glycosylate TAP (Figure 3.3). With the exception of ribose-5-phosphate (R5P), ribulose, erythrose, and threose, all sugars tested can adopt both furanose and pyranose ring structures, albeit in different proportions. Four of the sugars tested contain acidic ionizable groups, and one contains a basic group (glucosamine). We previously reported that TAP reacts with ribose to form glycosides in the dry state at 35°C, and more slowly in aqueous solution at 5°C [1]. In the present study, the reactivity of TAP (at 1 M) with various sugars (at 1 M) was investigated in aqueous

solutions at pH 1 or pH 7 that were held at 85°C for 24 hours. This reaction temperature was selected following preliminary experiments which revealed that glycosidic bond formation between TAP and some sugars (primarily hexoses) in water gave good yields (greater than 10%) at 85°C after 24 hours, whereas higher temperatures caused some sugars to rapidly degrade. A 1:1 molar ratio of sugar and TAP was selected to favor singly-glycosylated TAP, though previous work indicated single ribosylation of TAP at even higher ratios of ribose [1]. Reactions at pH 7 were investigated because neutral pH is more plausibly prebiotic than strongly acidic or basic conditions [10], which require invoking special geochemical environments. Reactions at pH 1 were carried out to promote *C*-glycoside formation based on preliminary studies that showed enhanced production of the β-*C*-ribofuranoside when TAP reacts with R5P at low pH. Solution state reactions were selected because dry-state model prebiotic reactions can exhibit substantial variations in yields [11], which could complicate the comparison of yields between sugars. Even with these unoptimized conditions, analyses of reaction products by UV-LC/MS and ^{1}H NMR spectroscopy indicate that all sugars tested react with TAP to form glycosides. As an example of one such analysis, UV-monitored LC/MS chromatograms and ^{1}H NMR spectra of the crude products from the reactions of TAP with glucose are shown in Figure 3.4.

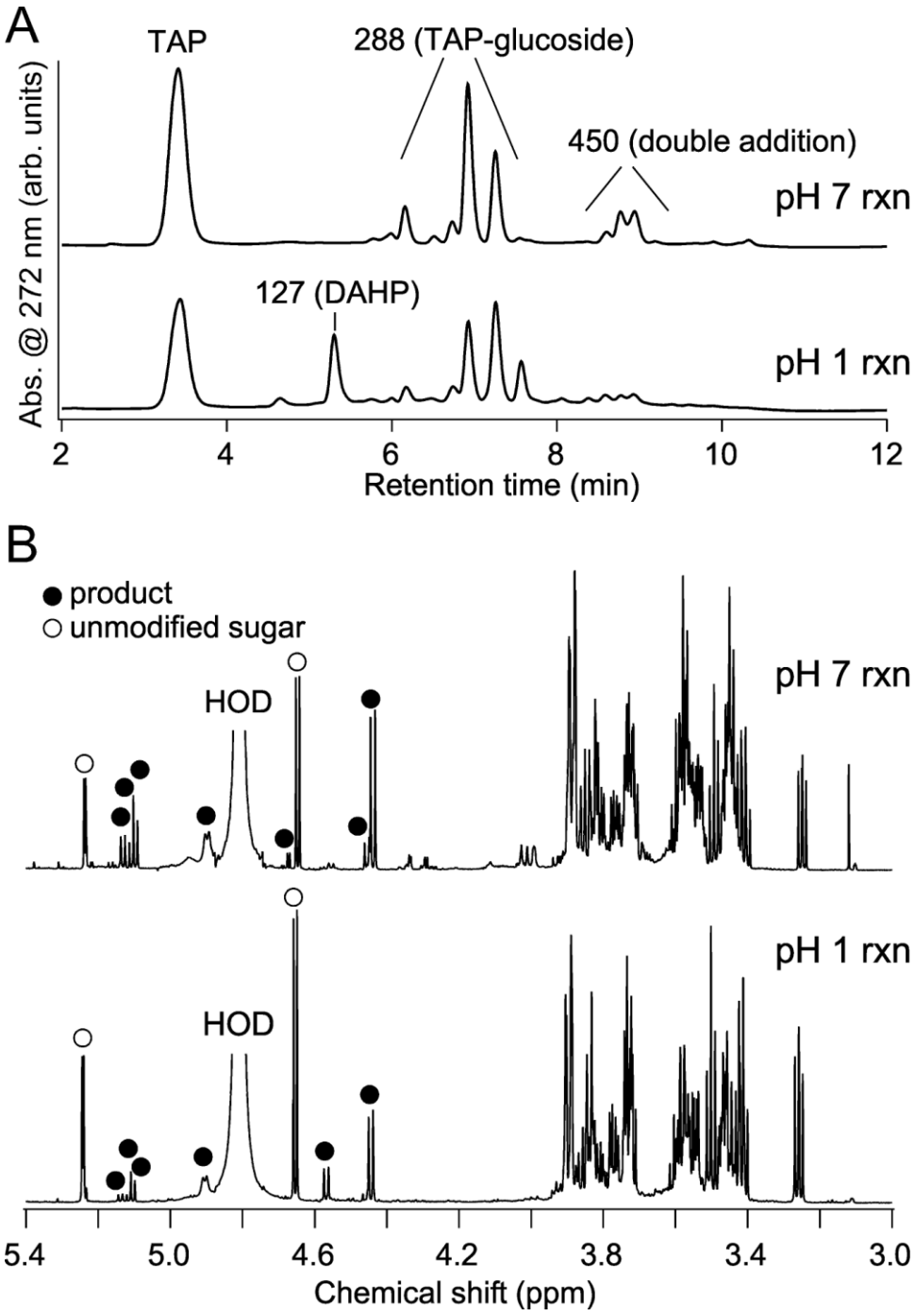


Figure 3.4. Example analyses of crude products resulting from the reaction of TAP (1 M) with glucose (1 M) at 85°C for 24 hours at pH 1 and pH 7. A. UV-LC/MS chromatograms monitored by absorbance at 272 nm, with m/z values associated with peaks as indicated. DAHP is diaminohydroxypyrimidine, a hydrolysis product of TAP. B. ^{1}H NMR spectra of the same crude products for which chromatograms are shown in A. Spectra were acquired in D_2O at 20°C. Adapted with permission from [2].

3.4.2 Detailed Analysis of Glucose, N-Acetylglucosamine, and Glucose-6-Phosphate Glycosides

Because previous studies of TAP glycoside formation were performed with ribose (a pentose) and our initial analysis revealed strong evidence of glycoside formation with glucose (a hexose), we decided to perform a detailed structural analysis of the TAP-glucose reaction products. Additionally, to assess the impact of sugar modifications on TAP glycosylation, we performed the same structural analysis of products formed in TAP reactions with glucose-6-phosphate (D-Glc6P) and *N*-acetylglucosamine (D-GlcNAc).

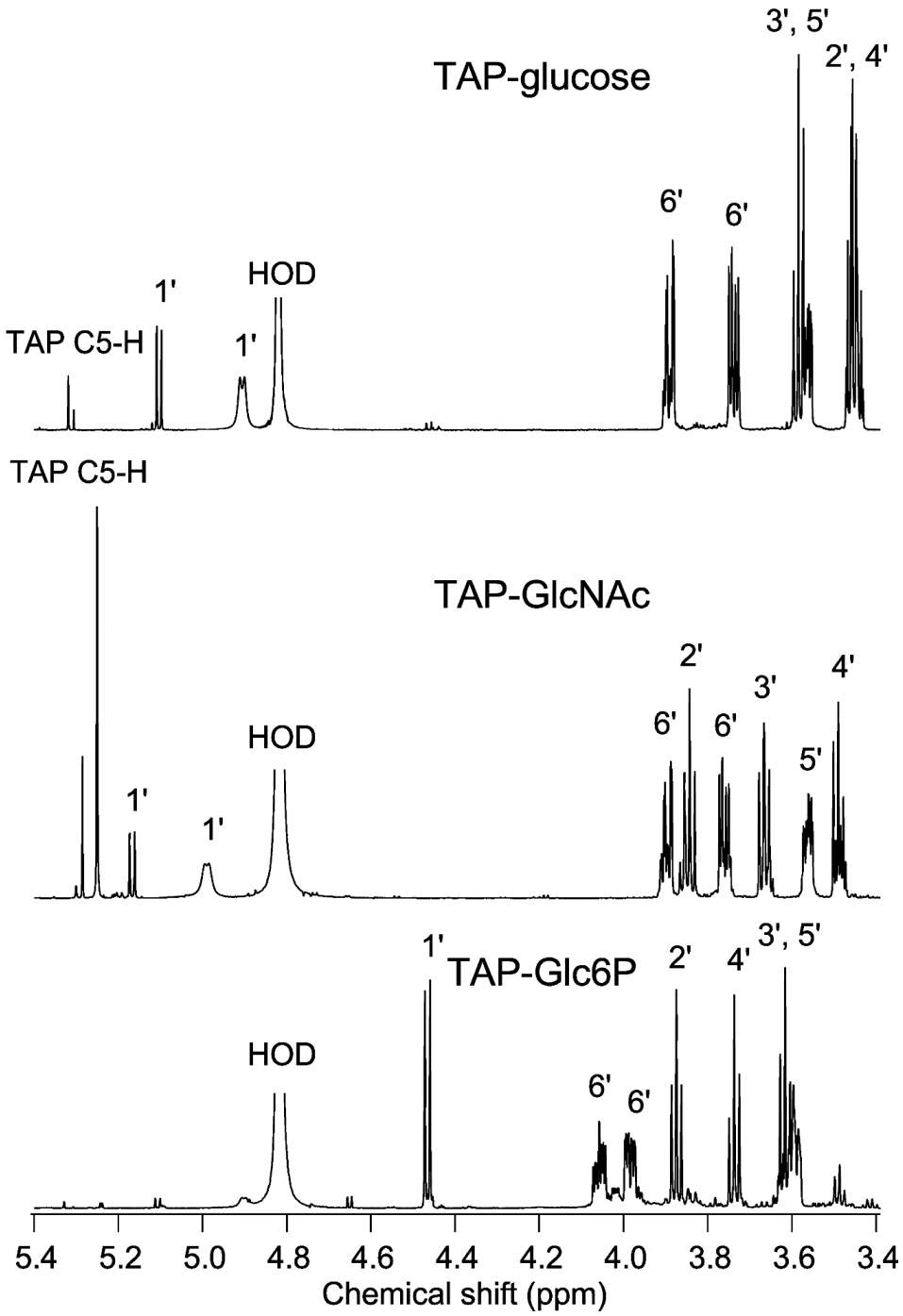


Figure 3.5. ^{1}H NMR spectra of glycosides isolated from the products of the TAP-glucose, TAP-GlcNAc, and TAP-Glc6P reactions performed at pH 1. Spectra were acquired at 20°C with D_2O as the solvent. Resonance assignments are based on ^{1}H COSY spectra. Procedures used for purification are described in § 3.2.4. Adapted with permission from [2].

The main products resulting from the reactions of TAP with glucose, D-Glc6P, and D-GlcNAc at pH 1 were isolated and the information gained from their structure determination was used to characterize the products of the pH 7 reactions, as well as additional products in the pH 1 reactions that were not isolated. Several lines of evidence, based on UV-LC/MS and ^{1}H, ^{13}C NMR spectroscopy, were used to confirm the structures of the purified glycosides. ^{1}H NMR spectra of the isolated products showed that the reactions of TAP with glucose and GlcNAc each produced two glycosides as major products that co-eluted when purified by preparative reverse phase chromatography (Figure 3.5). In contrast, the reaction of TAP with Glc6P produced only one major glycoside product that was isolable by cation exchange chromatography (Figure 3.5). Before analysis by NMR spectroscopy, the purified reaction products were deuterium-exchanged and dissolved in a basic D_2O phosphate buffer, pD 12, to avoid acid-catalyzed hydrolysis of glycosides. The spectra of the TAP-glucose products and the TAP-GlcNAc products both showed two sharp singlets around 5.2 ppm that slowly decrease in intensity over time. The parent heterocycle TAP shows a similar diminishing resonance in D_2O, as the aromatic proton at position C5 exchanges slowly in protic solvents. Thus, the observation of exchangeable protons in the ^{1}H NMR spectra of the TAP-glucose and TAP-GlcNAc isolated products indicate that the two products in both samples are *N*-glycosides. No such diminishing resonance was observed in the spectrum of the isolated TAP-Glc6P product, consistent with a *C*-glycoside of TAP (see below).

1D ROE analysis of the anomeric (H1′) signals of the isolated products of TAP-glucose, TAP-GlcNAc, and TAP-Glc6P reactions indicate that all five isolated glycosides are β-pyranosides, consistent with H1′-H2′ ^{3}J coupling constants of around 9 Hz in all cases

(Figure 3.5). Due to the symmetry of TAP, substitutions at the amino groups of positions 4 and 6 are equivalent; therefore, only two β-*N*-pyranosides are possible for both TAP-glucose and TAP-GlcNAc (substituted at position 2 or position 4/6 of TAP), which are the two observed (see Figure 3.6).

Confirmation of the glycoside structures was provided by 2D heteronuclear NMR spectra. HMBC analysis of the single isolated TAP-Glc6P glycoside revealed coupling of the anomeric proton to a relatively upfield signal at 84 ppm not assigned to any sugar carbon atom by HSQC. This resonance is also close in chemical shift to C5 of the parent TAP heterocycle. These through-bond correlations and ^{13}C chemical shift provide further evidence that the isolated TAP-Glc6P product is a *C*-glycoside. Additionally, the upfield chemical shift of the anomeric proton of TAP-Glc6P (i.e., 4.46 ppm) is consistent with *C*-substitution of TAP when compared to other *C*-glycosides [1], and dissimilar to the more downfield chemical shifts of *N*-glycosides (i.e., TAP-glucose, 4.9 ppm and 5.1 ppm; TAP-GlcNAc, 5.0 and 5.2 ppm).

While only *N*-glycosides were isolated from the reactions of TAP with glucose and GlcNAc, resonances are present in the ^{1}H NMR spectra of the crude reaction mixtures that, due to the more upfield chemical shift of their putative anomeric resonances, suggested the presence of *C*-glycosides. Similarly, although a *C*-glycoside was isolated from the reaction of TAP with Glc6P, more downfield resonances in the ^{1}H NMR spectrum of the TAP-Glc6P crude reaction mixture suggested that *N*-glycosides were also formed.

To obtain additional structural information about these unpurified TAP glycosides, the crude reaction products of the pH 1 and pH 7 reactions were spiked with the TAP

glycosides isolated from their corresponding pH 1 reactions. This process allowed assignment of the purified glycoside anomeric resonances within the crude-products spectra. Interestingly, all of the glycosides isolated from the pH 1 reactions were also present in the corresponding pH 7 reactions. Spiking of these samples with the reactant sugar also allowed identification of unreacted sugar resonances.

The presumed anomeric resonances in the NMR spectra of the crude products that were not assigned to a purified glycoside or parent sugar were analyzed by acquiring 1D ROE spectra. As mentioned above, in both the pH 1 and pH 7 TAP-glucose reactions, one major product resonance was present upfield of the water peak, indicative of a *C*-glycoside (Figure 3.4). 1D ROE analysis of this signal produced a spectrum similar to the 1D ROE obtained for the purified TAP-Glc6P product (determined to be a *C*-glycoside), with through-space interactions from the anomeric proton (1′) to signals matching those of the sugar 3′ and 5′ protons, and pronounced TOCSY transfer to the 2′ proton, suggesting a β-pyranoside. In these spectra, the 2′ proton resonance is also more downfield (at 3.86 ppm) than in the spectra of purified TAP-glucose (determined to be *N*-glycosides).

The ^{1}H NMR spectrum of the TAP-GlcNAc pH 7 reaction products did not exhibit any anomeric proton resonances upfield of the water peak, suggesting no appreciable *C*-glycoside formation. The ^{1}H NMR spectrum of the TAP-GlcNAc products from the pH 1 reaction did exhibit a resonance with a chemical shift in the range expected for an anomeric proton of a *C*-glycoside. However, the 1D ROE transfers associated with this resonance were not indicative of a β-*C*-glycoside.

Additional anomeric resonances are observed in the ^{1}H NMR spectra of the crude reaction mixtures of TAP with Glc6P (from which a *C*-glycoside was isolated). In the pH 7 reaction, two of these resonances are located downfield of the water peak, consistent with the formation of *N*-glycosides. Furthermore, one peak is characteristically broad, and the other sharp, similar to the anomeric resonances of the *N*-glycosides isolated from the TAP-GlcNAc and TAP-glucose reactions. 1D ROE analysis of these TAP-Glc6P reaction products also exhibit magnetization transfers similar to those of the TAP-glucose and TAP-GlcNAc N-glycosides, reinforcing the *N*-glycoside assignment and also suggesting β-pyranosides for these unpurified products of the pH 7 TAP-Glc6P reaction. Similar resonances are also present in the spectra of the pH 1 TAP-Glc6P reaction mixture, but were too weak for 1D ROE analysis.

Taken together, our analyses of purified glycosides and crude reaction products indicate that the reactions of TAP with glucose at either pH 1 or 7 produce the two possible β-*N*-pyranosides and the one possible β-*C*-pyranoside (Figure 3.6). The reactions of TAP with GlcNAc at either pH 1 or 7 apparently produce only the two possible β-*N*-pyranosides. The reactions of TAP with Glc6P also produce the β-*N*-pyranoside, but predominantly the β-*C*-pyranoside, especially at pH 1 (Figure 3.6).

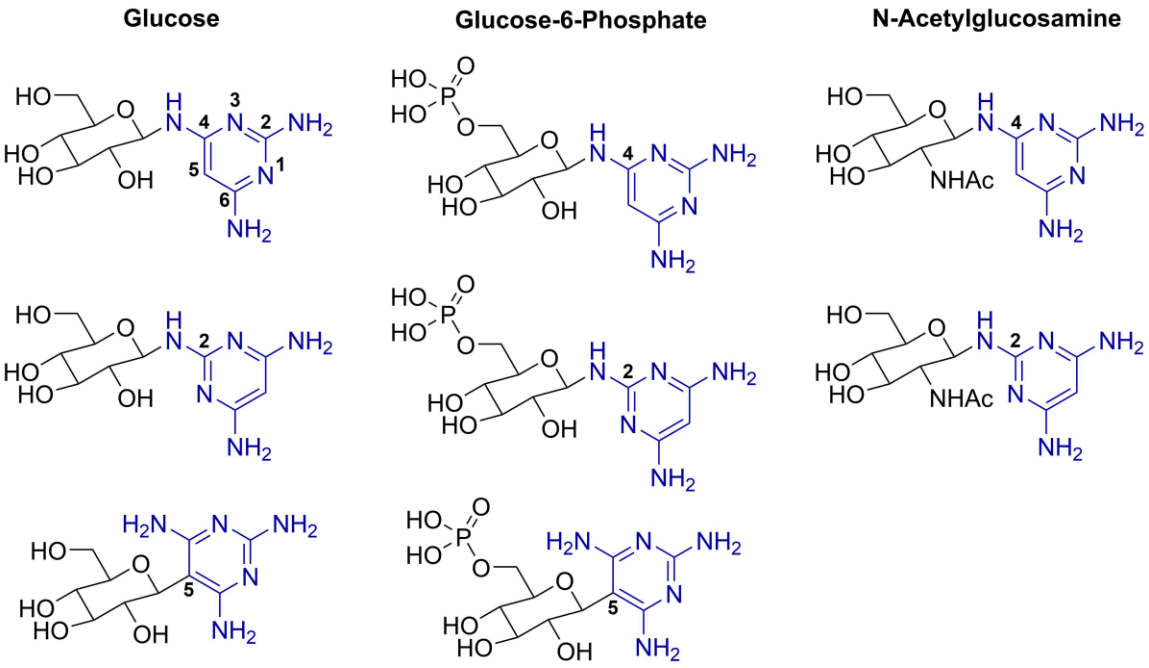


Figure 3.6. Chemical structures of the TAP glycosides identified among the products of the TAP-glucose, TAP-GlcNAc, and TAP-Glc6P reactions. Note that all glycosides are of the β-pyranose form. The three possible glycosides of this form were identified among the TAP-glucose and TAP-Glc6P reaction products. The *C*-glycoside was not detected among the TAP-GlcNAc reaction products. Adapted with permission from [2].

It is intriguing that all eight of the glycosides identified among the reaction products of TAP with glucose, GlcNAc, and Glc6P are of the β-pyranose form. Previous studies of model proto-nucleobase reactions with ribose have also revealed a preference for the formation of the β-anomer over the α-anomer [11, 12].

3.4.3 The Mechanism of Glycosylation of TAP

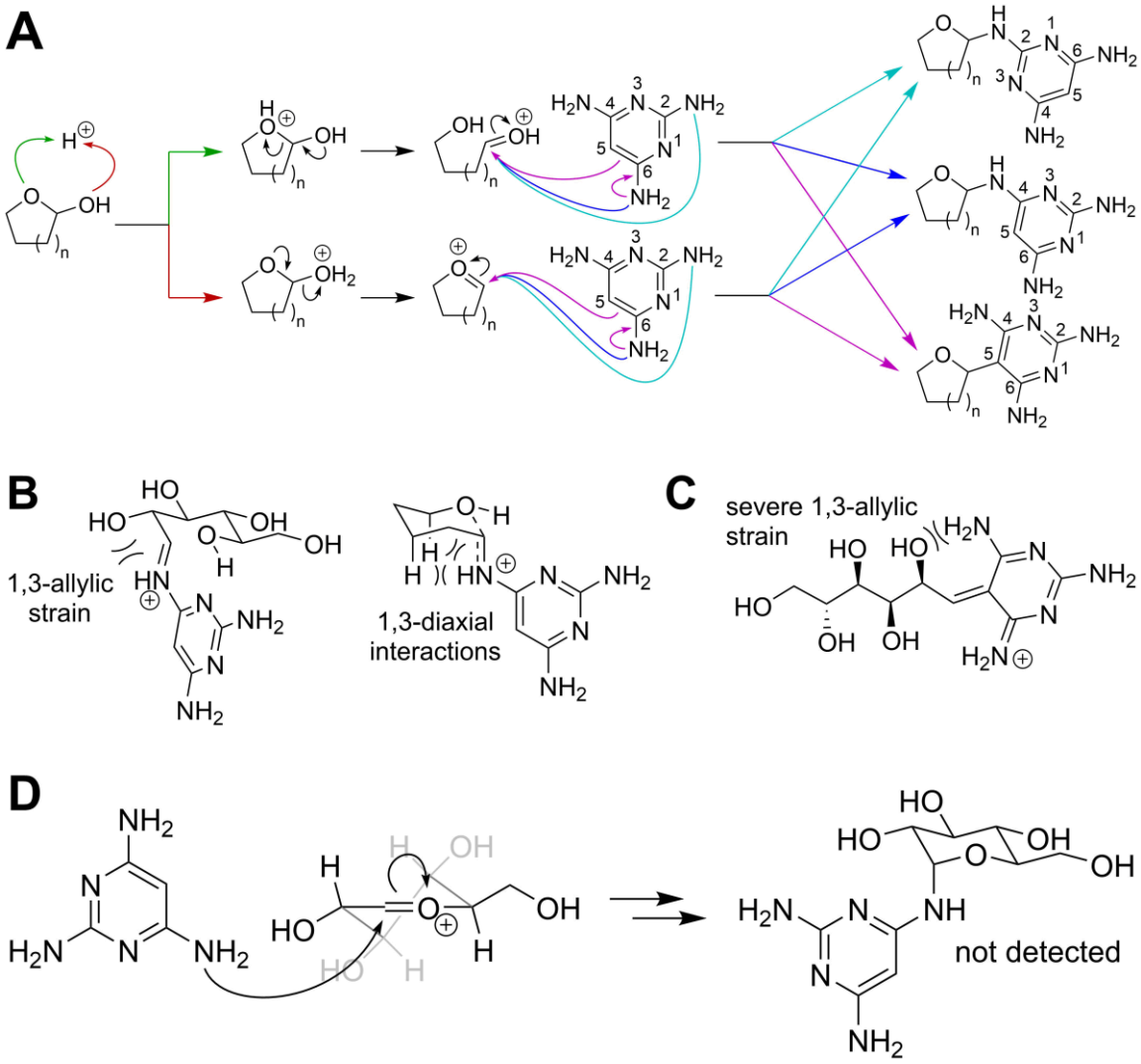


Figure 3.7. The possible acid-catalyzed mechanisms of glycosylation of TAP. A. Electrophilic activation of a sugar may occur through endocyclic protonation, followed elimination of the hydroxyl group and subsequent attack of the aldehyde oxocarbenium moiety (green path), or through exocyclic protonation, followed by elimination of water and attack of the cyclic oxocarbenium moiety (red path). B. For glucose derivatives, *N*-glycosylation by attack of a protonated aldehyde will disfavor the formation of α-pyranosides by 1,3-allylic strain in the intermediate, and corresponding 1,3-diaxial interactions in the transition state, leading to the product. C. *C*-Glycosylation of TAP by attack of the protonated aldehyde requires the formation of a Knoevenagel-like intermediate with severe 1,3-allylic strain. D. *C*-Glycosylation of TAP by attack of the cyclic oxocarbenium moiety would kinetically favor the formation of α-pyranosides, which are not observed. Adapted with permission from [2].

The possible mechanisms of TAP glycosylation are shown in Figure 3.7A. Glycosylation in water is expected to be acid-catalyzed, and may proceed either by protonation of the endocyclic oxygen atom, followed by endocyclic C-O bond cleavage to give a protonated aldehyde electrophile (as an acyclic-form sugar, green path), or by protonation of the exocyclic anomeric hydroxyl group, followed by exocyclic C-O bond cleavage to give an oxocarbenium electrophile (as a cyclic-form sugar, red path). Previous studies on aqueous glycosylation of heteroaromatic compounds [13] supported the endocyclic (green) mechanism that proceeds by attack of the nucleophilic heterocycle on the protonated aldehyde of an acyclic-form sugar. However, for pyranosides (glucose presents mainly as α- and β-pyranoses in water), solvolysis generally follows the exocyclic (red) oxocarbenium mechanism, suggesting that the reverse reaction, glycosylation, also follows this mechanism [14].

The dominant mechanism in the glycosylation of TAP by glucose and its derivatives is unclear. The acyclic path is reasonable for the formation of β-*N*-pyranosides, first proceeding through Schiff base formation with an exocyclic amine of TAP, followed by attack of the 5′ hydroxyl group on the protonated Schiff base to form the glycosylamine. Because of allylic strain in the Schiff base intermediate, and corresponding diaxial strain in the transition state leading to the α-pyranoside, this product is not observed (Figure 3.7B). However, for β-*C*-pyranoside formation through an acyclic sugar, a Knoevenagel condensation-type intermediate must be formed which would have severe 1,3-allylic strain (Figure 3.7C). The formation of β-*C*-pyranosides is not ameliorated through the cyclic path, which, through consideration of the most stable half-chair conformation of the oxocarbenium species, is expected to give α-pyranosides (Figure 3.7D), which are not

observed. A more quantitative computational analysis of the possible mechanisms may demonstrate which path is more likely.

Although anchimeric assistance by the acetamido group of GlcNAc is expected to enhance β-substitution, the formation of this more stabilized cation may not provide sufficient electrophilicity to incite electrophilic aromatic substitution of TAP, thus preventing the formation of the β-*C*-pyranoside.

3.4.4 Survey of TAP Glycoside Formation with Other Sugars

Our confirmation of glycoside formation by TAP with glucose and two glucose derivatives in the current study, and with ribose in a previous study [1], suggest that TAP glycosylation may be possible with a wide variety of sugars. Indeed, UV-LC/MS and ^{1}H NMR spectroscopic analyses of crude reaction mixtures indicate that all of the sugars shown in Figure 3.3 produce TAP glycosides at pH 7. These reactions were carried out with the same conditions used for the TAP-glucose/GlcNAc/Glc6P reactions (i.e., 1 M TAP, 1 M sugar, 24 h, 85°C, pH 1 or pH 7).

The thorough characterization of the TAP glucosides enables the preliminary assessment of the formation of a variety of glycosides from TAP by analysis of the UV-LC/MS chromatograms and 1D ^{1}H NMR spectra of crude reaction mixtures. Specifically, the number of new anomeric resonances observed in the ^{1}H NMR spectra and the number of product peaks observed by LC/MS are reported as an estimate of the number of unique glycosides formed by TAP with a given sugar (Table 3.1). The integrated intensities of these ^{1}H resonances (Table 1) and UV-monitored chromatogram peaks were likewise used to estimate glycoside yields. For comparison, the TAP-glucose reaction produced

glycosides in 31% and 44% yield of total remaining TAP/glucose for the pH 1 and pH 7 reactions, respectively. These yields are typical of most of the sugars tested, which generally produce TAP glycosides with estimated yields that range from 20 to 50% (Table 3.1). Integration of UV absorption peaks after HPLC separation was used as a second means to estimate glycoside yields, with coincident mass spectral analysis used to identify peaks corresponding to TAP glycosylated by a single sugar. Although yields based on UV-HPLC absorption peak integration are higher for most sugars compared to yields obtained by ^{1}H NMR resonance peak integration, there is a positive correlation between yields measured by the two techniques (Figure 3.8) Thus, while these yields may be affected by formation of insoluble products, these values provide reliable relative propensities for these seventeen sugars to glycosylate TAP.

Table 3.1. Estimated Number of Glycosides and Yields for the Reaction of TAP with Various Sugars[A]

Sugar	pH	Anomeric ^{1}H	Product Peaks (LC/MS)	Product Yields (est.)[B]	Unreacted Sugar[B]
Hexoses					
Glucose	1	8	6	31%	43%
	7	10	8	44%	29%
GlcNAc	1	10	8	28%	63%
	7	4	2	5%	28%
Glc6P	1	4	2	27%	41%
	7	5	2	22%	25%
Fructose[C]	1	2	2	2%	NA
	7	3	3	12%	NA
Galactose	1	8	6	35%	34%
	7	7	5	48%	19%
Galacturonic Acid	1	12	12	40%	0%
	7	8	6	53%	14%
Glucosamine	1	6	4	10%	40%
	7	2	2	15%	0%
Glucuronic Acid	1	10	8	42%	18%
	7	5	3	24%	11%
Mannose	1	3	1	20%	30%
	7	5	3	13%	18%
Pentoses					
Arabinose	1	7	4	61%	43%
	7	7	5	55%	13%
Lyxose	1	4	2	20%	25%
	7	6	4	40%	11%
Ribose	1	11	7	28%	24%
	7	8	7	31%	3%
R5P	1	4	4	11%	0%
	7	4	4	8%	0%
Ribulose[C]	1	3	3	3%	NA
	7	11	11	31%	NA
Xylose	1	7	5	27%	28%
	7	9	7	34%	9%
Tetroses					
Erythrose[D]	1	1	3	<1%	0%
	7	3	3	17%	0%
Threose[D]	1	7	3	3%	0%
	7	2	3	2%	0%

A. Unique product species estimated by number of non-sugar resonances in anomeric region of ^{1}H NMR spectra and product peaks in LC/MS chromatograms. Estimated product yields and amounts of unreacted sugars provided are based on integrated intensities of anomeric resonances relative to an internal standard. Not all product anomeric signals identified are necessarily TAP glycosides. Estimated product yields based on integration of UV-monitored LC-MS chromatograms are provided in ESI. B. TSP was used as an internal concentration standard in NMR samples. The exclusion of insoluble products from solution-state NMR analysis can cause the estimated product yields and unreacted sugar concentrations to not always sum to 100% of the original sugar concentration. It is well known that sugar degradation can lead to insoluble "tars" [15], or isomerization to a different sugar, leading to products not closely associated with the parent sugar. C. Estimated yields for fructose and ribulose are based on integration of anomeric ^{1}H resonances, which would be limited to TAP glycosides formed with aldose sugars that result from fructose and ribulose isomerization, respectively. D. Dissolving threose or erythrose in the NMR buffer (pD 12) results in rapid degradation. Thus, the NMR samples used to determine amounts of unreacted sugar may give values with large errors for these two sugars. Adapted with permission from [2]/

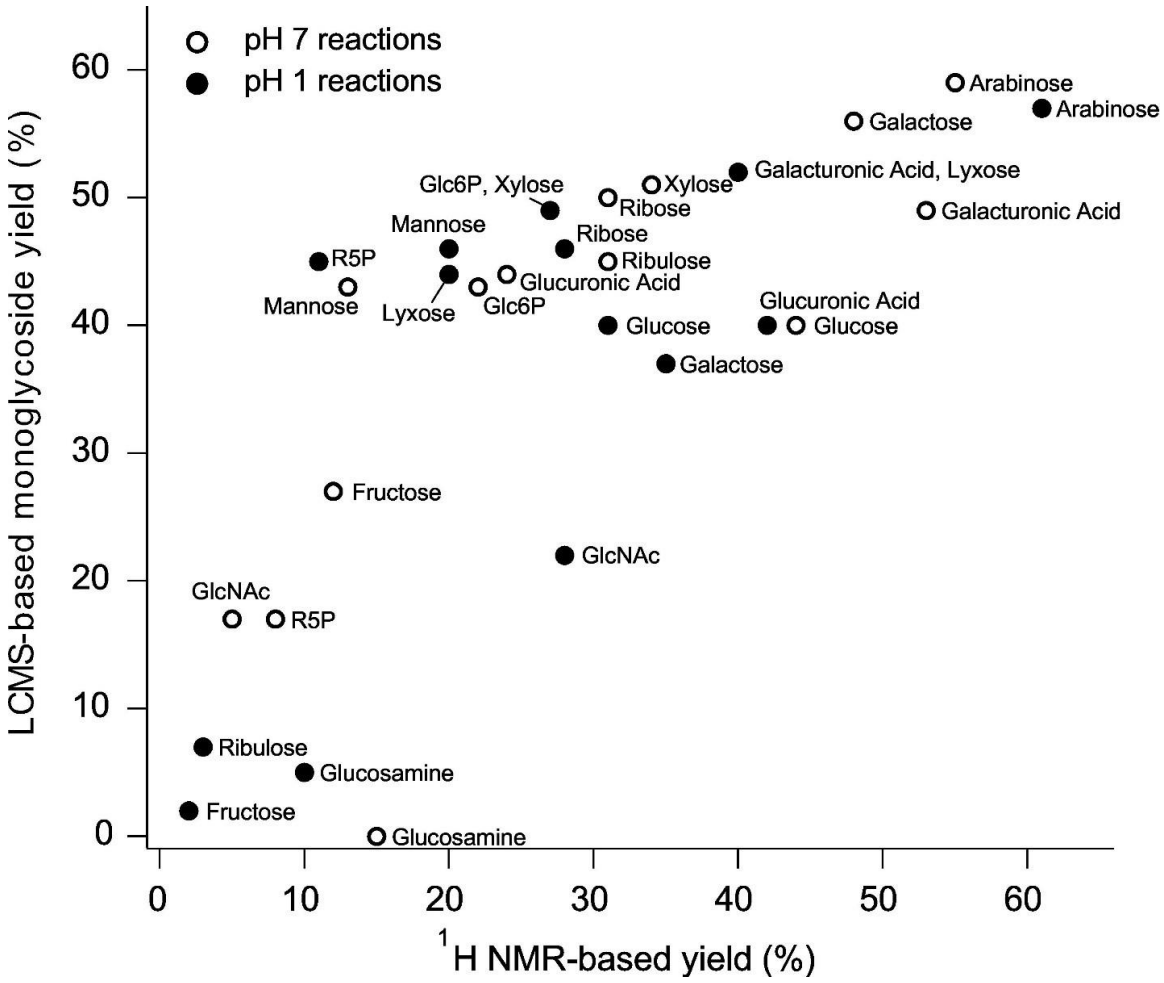


Figure 3.8. Plot of estimated TAP-glycoside yields based on integrated absorption at 272 nm of peaks observed in UV-LC/MS chromatograms with m/z values corresponding to single sugar conjugates of TAP versus estimated TAP-glycoside yields based on integration of new (i.e., excluding unreacted sugar) resonances in the anomeric regions of ^{1}H NMR spectra. UV-LC/MS-based yields are relative to total integrated intensity of peaks with absorption at 272 nm (i.e., total recovered TAP and TAP conjugates). ^{1}H NMR-based yields are determined by comparison of integrated intensities to the integrated intensities of the resonances of an internal standard of known concentration. Adapted with permission from [2].

Inspection of the estimated glycoside yields provided in Table 1 does not reveal an obvious pattern for either the pentoses or the hexoses. Nevertheless, the products of TAP incubation with a given sugar can, in some cases, be understood from the known properties of the sugar. For example, the reaction of TAP with ribose-5-phosphate (R5P) appears to not produce a noncanonical nucleotide. This may be due in part to the rapid degradation of

R5P, which has an estimated half-life of only one hour at 85°C [16]. Similarly, the propensity of two glucosamine molecules to react and spontaneously form 2,5-dihydropyrazines in a Maillard-like reaction may limit the formation of TAP glycosides with glucosamine [17]. The degradation rate of a sugar is, of course, only one of several factors expected to govern TAP-glycoside yields. The rate at which a given sugar glycosylates TAP is also expected to be an important determinant of glycoside yield, as glycoside formation greatly reduces (or even eliminates) the amount that an aldose sugar exists in equilibrium with its open-chain free aldehyde form, and the equilibrium amount, which strongly correlates with its degradation rate [16]. In this context, the low yields of TAP glycosylation by erythrose and threose could be, in part, due to these sugars being degradated more rapidly than other sugars (prior to protection by glycoside formation with TAP), as the fraction that these tetrose sugars exist in their open-chain aldehyde form is ca. 20-fold greater than that of ribose and ca. 400-fold greater than that of glucose [18, 19].

The reaction of TAP with uronic acids was complicated by the formation of low solubility salts (see § 2.6). Specifically, when TAP and galacturonic acid or glucuronic acid were dissolved in water, co-precipitates rapidly form, which can be re-solubilized by heating at 100°C with frequent and rapid stirring. Once solubilized, the components remain in solution at 85°C, pH 1 and pH 7, with the apparent formation of glycosides in relatively good yields (Table 1). For these samples in particular, the quantity of unreacted sugar reported in Table 1, which is based on ^{1}H resonance integration, may be significantly underestimated due to the reprecipitation of unreacted starting materials during NMR sample preparation.

Finally, characterization of the products formed by the reaction of TAP with fructose and ribulose by ^{1}H NMR spectroscopy is complex because glycosides formed with these ketoses lack the anomeric proton resonances that are useful for identifying glycosides formed with aldose sugars. Peaks in UV-LC/MS chromatograms with m/z values matching TAP hexosides and TAP pentosides with fructose and ribulose, respectively, were detected among the pH 7 reactions products, and to a lesser extent among the pH 1 reaction products. The ^{1}H NMR spectra of crude reaction products of TAP with fructose and ribulose *did* contain anomeric resonances, suggesting that these ketoses may have isomerized to aldoses. Isomerization of fructose can lead to glucose and mannose, while isomerization of ribulose can lead to ribose and arabinose [20-23], which may react with TAP to form glycosides. Thus, it is possible that, in pH 7 solutions of TAP with either fructose or ribulose, general base catalysis by TAP (pK_a = 7) may give the corresponding aldoses, which then react with TAP to form glycosides. Integration of the UV-HLPC peaks indicate similar yields of TAP-sugar conjugates as determined by integration of anomeric resonances in ^{1}H NMR spectra. Thus, although formation of TAP glycosides with ketose sugars cannot be ruled out, it appears that the aldose glycosides are formed in greater yields.

The formation of TAP-ribosides in reactions that start with ribulose as the only sugar is an intriguing observation from an origins-of-RNA perspective, as the chemical instability and lack of a robust prebiotic synthesis of ribose has been a long-standing conundrum for prebiotic chemists [16]. Benner and coworkers have argued that borate could have aided ribose formation and survival on the prebiotic Earth [24, 25]. However, the prebiotic existence of borate minerals has been questioned [26]. Ribulose, on the other hand, can be synthesized from dihydroxyfumarate and glyceraldehyde in reasonably good

yield (ca. 20%) [27]. This plausible prebiotic synthesis of ribulose, combined with our previous demonstration that TAP readily forms both *N*-glycosides and *C*-glycosides with ribose in good yields [1], and the current finding that ribulose will isomerize in the presence of TAP, presents an alternative possibility for a prebiotic origin of ribosides that involves the isomerization of a ketose sugar.

3.5 Formation of Supramolecular Assemblies with TAP Glucosides

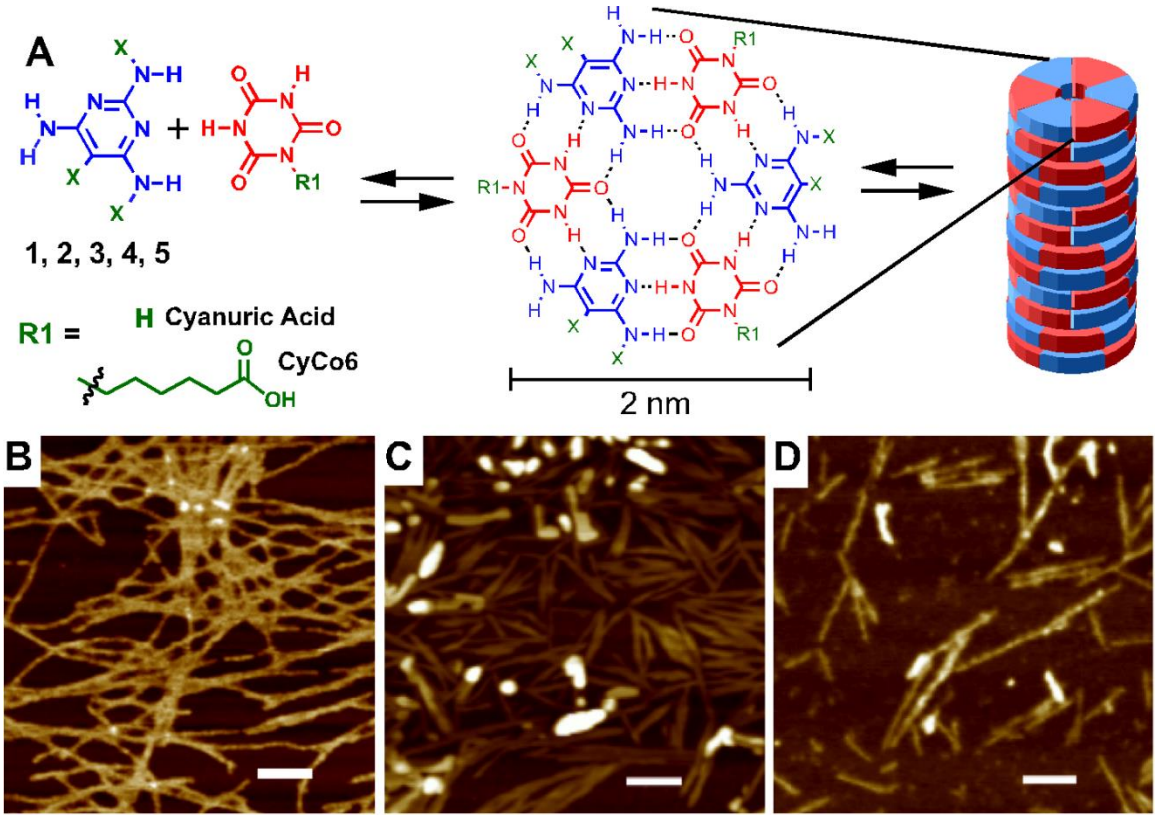


Figure 3.9. A. Chemical structure of TAP-cyanuric acid supramolecular assembly, with glycoside structures shown explicitly in Figure 3.6, the structure of cyanuric acid, CyCo6, and proposed hexad and stacked hexad assemblies. B. AFM images of assemblies formed by the crude products of the TAP-Glc6P reaction when mixed with cyanuric acid (50 mM total TAP/Glc6P, 50 mM CA, pH 7, and ca. 0.1 M NaCl from pH adjustment by addition of NaOH). Gelation was observed prior to deposition on mica surface. C. AFM image of assemblies formed by purified TAP-glucose glycosides (from pH 1 reaction) when mixed with CyCo6 (50 mM in total TAP species, 50 mM, CyCo6, pH 7, and ca. 0.3 M NaCl from pH adjustment with NaOH). No gelation or precipitation was observed prior to deposition. D. AFM images of purified TAP-GlcNAc glycosides formed when mixed with CyCo6 (50 mM total TAP species, 50 mM CyCo6, pH 7, and ca. 0.3 M NaCl from pH adjustment with NaOH). No precipitation or gelation was observed prior to deposition. Scale bar is 100 nm in all AFM images. Adapted with permission from [2].

We assessed the propensity for the glycosides formed by glucose and its derivatives to form supramolecular assemblies in water with cyanuric acid or a charged derivative of cyanuric acid (i.e., CyCo6,23 cyanuric acid substituted with hexanoic acid, Figure 3.9).

Such assemblies are relevant to our working hypothesis that noncovalent assemblies with Watson-Crick or Watson-Crick-like base pairs were critical for the prebiotic synthesis of the first nucleic acids [28, 29]. Such assemblies could have facilitated proto-RNA synthesis by locally organizing, concentrating, and segregating the bases of proto-RNA from other heterocycles that do not have the ability to base pair. As illustrated in Figure 3.9A, TAP and TAP with pendant groups (e.g., sugars) can form hexads with cyanuric acid and cyanuric acid derivatives that assemble into coaxial stacks in water.

With a peripheral ionizable phosphate group, we expected the TAP-Glc6P glycosides to have the greatest propensity to form soluble assemblies, as an electrostatic charge is important for maintaining the solubility of stacked TAP-cyanuric acid hexads [3]. Indeed, mixing the crude TAP-Glc6P reaction products with cyanuric acid did result in the appearance of the expected assemblies, as revealed by AFM imaging (Figure 3.9B). In contrast, and unexpectedly, the purified TAP-Glc6P glycoside formed insoluble precipitates when mixed with cyanuric acid (reminiscent of the precipitation of TAP-GlcA with cyanuric acid; see § 2.6).

Linear assemblies are observed by AFM for the purified TAP-glucose glycosides when mixed with CyCo6 (Figure 3.9C). In contrast, insoluble precipitates of irregular structure were observed when the purified products of TAP-glucose reaction were mixed with CyCo6. Similarly, linear structures were formed when purified TAP-GlcNAc was mixed with CyCo6 (Figure 3.9D), but mixing crude TAP-GlcNAc with CyCo6 did not. These results suggest that the steric determinants of TAP-cyanuric acid assembly into linear structures in water are not trivial. Work is on-going to understand the effects of the

crude versus the purified products, as well as sugar species, on the propensity for TAP glycosides to form soluble assemblies.

3.6 Conclusions

Scenarios for the prebiotic origin of nucleic acids often assert (explicitly or implicitly) some mechanism for the selective production of β-ribofuranosides over other glycosides. However, a robust and prebiotically plausible synthesis of the canonical β-ribofuranosides is still lacking, and even the selective synthesis of ribose remains a challenge. Sugar formation is inherently non-selective, as the products of reactions that produce sugars can also act as substrates for further homologation. Therefore, it is important to assess whether or not a reactive, and potentially information-transferring heterocycle, such as TAP, has the ability to react with a variety of sugars in the same manner that it does with ribose. Even under the non-optimized reaction conditions used in this study, TAP apparently reacts with all sugars surveyed, at pH 7, to give glycosides in moderate to good yields. These results call into question the assumption that prebiotic nucleotide selection occurred at the monomer level, since, in a prebiotically realistic scenario, reactive heterocycles would have not likely discriminated between the different sugars of a complex mixture, and therefore a variety of glycosides would be present. Organic chemists have demonstrated that some sugars are able to substitute for ribose in the backbone of RNA without disrupting the ability to form duplexes with Watson–Crick base pairs [30]. Here we have also shown that different glycosides of TAP show different propensities towards self-assembly with a pairing heterocycle (e.g., cyanuric acid). These observations suggest that if there was a wide variety of nucleosides on the prebiotic Earth, higher-order structures may have influenced proto-RNA monomer selection.

3.7 References

[1] Chen, M.C., et al., Spontaneous prebiotic formation of a β-ribofuranoside that self-assembles with a complementary heterocycle. Journal of the American Chemical Society, 2014. 136(15): p. 5640-5646.

[2] Fialho, D.M., et al., Glycosylation of a model proto-RNA nucleobase with non-ribose sugars: implications for the prebiotic synthesis of nucleosides. Organic & Biomolecular Chemistry, 2018. 16: p. 1263-1271.

[3] Cafferty, B.J., et al., Efficient self-assembly in water of long noncovalent polymers by nucleobase analogues. Journal of the American Chemical Society, 2013. 135(7): p. 2447-2450.

[4] Joshi, A.A., S.S. Narkhede, and C.L. Viswanathan, Design, synthesis and evaluation of 5-substituted amino-2,4-diamino-8-chloropyrimido-[4,5-b]quinolines as novel antimalarials. Bioorg Med Chem Lett, 2005. 15(1): p. 73-6.

[5] Rakotondradany, F., et al., Hydrogen-bond self-assembly of DNA-analogues into hexameric rosettes. Chemical Communications, 2005(43): p. 5441-5443.

[6] Becker, S., et al., A high-yielding, strictly regioselective prebiotic purine nucleoside formation pathway. Science, 2016. 352(6287): p. 833-836.

[7] Becker, S., et al., Wet-dry cycles enable the parallel origin of canonical and non-canonical nucleosides by continuous synthesis. Nature Communications, 2018. 9(1): p. 163.

[8] Trinks, U.P., Zur Chemie der Aminopyrimidine, in Dipl. Chem. ETH. 1987, ETH Zürich: Zürich.

[9] Decker, P., P. Schweer, and R. Pohlmann, Identification of formose sugars, presumable prebiotic metabolites, using capillary gas chromatography/gas chromatography-mass spectroscopy of n-butoxime trifluoroacetates on OV-225. J. Chromatogr., 1982. 244: p. 281-291.

[10] Raven, J. and K. Skene, Chemistry of the early oceans: the environment of early life, in Evolution on Planet Earth: The Impact of the Physical Environment, L.J. Rothschild and A.M. Lister, Editors. 2003, Academic Press: New York. p. 55-64.

[11] Bean, H.D., et al., Formation of a β-pyrimidine nucleoside by a free pyrimidine base and ribose in a plausible prebiotic reaction. J. Am. Chem. Soc., 2007. 129: p. 9556-9557

[12] Kolb, V.M., J.P. Dworkin, and S.L. Miller, Alternative bases in the RNA world: The prebiotic synthesis of urazole and its ribosides. J. Mol. Evol., 1994. 38: p. 549-557.

[13] Dworkin, J.P. and S.L. Miller, A kinetic estimate of the free aldehyde content of aldoses. Carbohydr. Res., 2000. 39: p. 359-365.

[14] Horenstein, N.A., Mechanisms for nucleophilic aliphatic substitution at glycosides. Adv. Phys. Org. Chem., 2006. 41: p. 275-314.

[15] Benner, S.A., H.-J. Kim, and M.A. Carrigan, Asphalt, water, and the prebiotic synthesis of ribose, ribonucleosides, and RNA. Acc. Chem. Res., 2012. 45: p. 2025-2034.

[16] Larralde, R., M.P. Robertson, and S.L. Miller, Rates of decomposition of ribose and other sugars: implications for chemical evolution. Proc Natl Acad Sci U S A, 1995. 92(18): p. 8158-60.

[17] Hrynets, Y., M. Ndagijimana, and M. Betti, Studies on the Formation of Maillard and Caramelization Products from Glucosamine Incubated at 37 °C. Journal of Agricultural and Food Chemistry, 2015. 63(27): p. 6249-6261.

[18] Snyder, J.R. and A.S. Serianni, Deoxygenated and Alkylated Furanoses – Thorpe-Ingold Effects on Tautomeric Equilibria and Rates of Anomerization. Carbohydrate Research, 1991. 210: p. 21-38.

[19] Angyal, S.J., The composition of reducing sugars in solution. Adv. Carbohydr. Chem. Biochem., 1984. 42: p. 15-68.

[20] Takasaki, Y., Kinetic and Equilibrium Studies on d-Glucose-d-Fructose Isomerization Catalyzed by Glucose Isomerase from Streptomyces sp. Agricultural and Biological Chemistry, 1967. 31(3): p. 309-313.

[21] Takasaki, Y., Kinetic and Equilibrium Studies on d-Mannose-d-Fructose Isomerization Catalyzed by Mannose Isomerase from Streptomyces aerocolorigenes. Agricultural and Biological Chemistry, 1967. 31(4): p. 435-440.

[22] Tewari, Y.B. and R.N. Goldberg, An investigation of the equilibria between aqueous ribose, ribulose, and arabinose. Biophysical Chemistry, 1985. 22(3): p. 197-204.

[23] Stankovic, L. and J. Konigstein, ISOMERIZATION OF PENTOSES AND 2-PENTULOSES BY INORGANIC PHOSPHATES. Chemicke Zvesti, 1979. 33(3): p. 397-403.

[24] Kim, H.J., et al., Synthesis of carbohydrates in mineral-guided prebiotic cycles. Journal of the American Chemical Society, 2011. 133(24): p. 9457-9468.

[25] Ricardo, A., et al., Borate minerals stabilize ribose. Science, 2004. 303(5655): p. 196-196.

[26] Grew, E.S., J.L. Bada, and R.M. Hazen, Borate minerals and origin of the RNA world. Origins of Life and Evolution of Biospheres, 2011. 41(4): p. 307-316.

[27] Sagi, V.N., et al., Exploratory experiments on the chemistry of the "glyoxylate scenario": Formation of ketosugars from dihydroxyfumarate. Journal of the American Chemical Society, 2012. 134(7): p. 3577-3589.

[28] Hud, N.V., Searching for lost nucleotides of the pre-RNA World with a self-refining model of early Earth. Nature Communications, 2018. 9(1): p. 5171.

[29] Hud, N.V. and F.A.L. Anet, Intercalation-mediated synthesis and replication: a new approach to the origin of life. Journal of Theoretical Biology, 2000. 205(4): p. 543-562.

[30] Eschenmoser, A., Searching for nucleic acid alternatives. Chimia, 2005. 59(11): p. 836-850.

CHAPTER 4. THE REACTIONS OF CANONICAL AND NONCANONICAL NUCLEOBASES WITH MODEL PREBIOTIC ELECTROPHILES

4.1 Introduction

The reactions of noncanonical nucleobases with sugars including, but not limited to, ribose and its derivatives, is an exercise is agnosticism: the knowledge of the structure of *extant* informational polymers (i.e., the nucleic acids) should not bias our investigations into the *prebiotic* formation of informational polymers [1]. While maintaining the stance that the first informational, self-replicating polymers used nitrogenous heterocyclic compounds capable of Watson-Crick-like hydrogen-bonding recognition for information storage and transfer [2, 3], this agnosticism can be made more sincere by investigating the general trends in reactivity of several heterocyclic compounds with a variety of types of electrophiles that may have been present on the early Earth. In doing so, we gain information about what types of linkages could possibly have formed between heterocyclic compounds and candidate proto-nucleic acid backbone components, which furthers constrains the enormous chemical space of possible proto-nucleic acids.

Nucleobases and their chemical relatives are, in general, nucleophilic. When a purine or pyrimidine is substituted with amino groups, these substituents either act as electron-donating groups, enhancing the nucleophilicity of endocyclic sites (as is the case for the C5 position of TAP; see § 2.3.2), or act as nucleophiles themselves. When a purine or pyrimidine is *formally* substituted with hydroxyl groups, the keto tautomer dominates,

and the imidic endocyclic sites become mildly acidic (as does the methylene group at position 5 of barbituric acid; see § 2.3). Upon deprotonation, these imidate (or enolate) sites become nucleophilic. The same principles apply to other classes of heterocycles, such as triazines (e.g., melamine and cyanuric acid), triazoles (e.g., urazole), and imidazoles (e.g., hydantoin). Counter-examples of this principle of nucleophilicity exist, but are rare. Alloxan, the fully oxidized form of pyrimidine (with oxo substituents at positions 2,4,5, and 6), acts as an electrophile at position 5, and, to a lesser extent, at position 4, most notably with 1,2-diamino-substituted systems to form flavin derivatives [4]. Substitutions of nitrogenous heterocycles with electron-withdrawing groups are also possible, but are typically not detected in model prebiotic reactions or in meteoritic samples [5, 6].

Because nitrogenous heterocyclic compounds are usually nucleophilic, it follows that the means by which they are appended to proto-nucleic acid backbones must be through some electrophilic moiety. Therefore, a survey of reactivity between model proto-nucleic acid nucleobases (canonical or noncanonical) and model prebiotic electrophiles should give insight into what types of proto-nucleic acids are prebiotically feasible. This immediately raises an important question: what qualifies a class of electrophiles as plausibly prebiotic?

To answer this question, it is perhaps easiest to start with criteria that immediately disqualify a candidate electrophile. Electrophiles that are profoundly sensitive to hydrolysis, or are formed from reagents that are profoundly sensitive to hydrolysis, such as acyl chlorides, are almost certainly not prebiotically viable. This might also seem to apply to carboxylic acid anhydrides; however, anhydrides may be formed from dry-down reactions of carboxylic acids at elevated temperatures. If a solid residue containing

anhydrides is quickly rehydrated with an aqueous solution containing a nucleobases that is a better nucleophile than water, it may be possible to capture the anhydride before hydrolysis. Therefore, the anhydride will be considered in this analysis.

Another important disqualifying criterion is ease of prebiotic formation. For example, alkyl halides, although incredibly useful in conventional organic synthesis, are probably not produced in significant quantities by prebiotic chemical processes (in fact, such compounds have been considered atmospheric biosignatures [7]); therefore, they will not be considered.

With these disqualifying criteria in mind, we can being to populate our set of candidate prebiotic electrophiles. Carbonyl electrophiles are commonly invoked in prebiotic syntheses. For example, aldehydes act as the substrates for both the Strecker synthesis of amino acids [8] and the sugar-forming formose reaction [9]. Carboyxlic acids are also common [10], but are typically poor electrophiles in water due to deprotonation to form relatively inert carboxylates. However, certain carboxylic acid derivatives may serve as effective electrophiles. Esters, for example, can be formed from dry-down reactions of carboxylic acids and alcohols [11]. Thioesters have also been commonly noted as important for the emergence of life [12]. Amides, as the most stable carboxylic acid derivatives, will not be considered. However, imides, which are less electron-rich, may react with sufficiently strong nucleophiles. Finally, Michael acceptors have often been invoked in prebiotic syntheses, for example, in the synthesis of certain amino acids [13] and nucleobases, such as the syntheses of aspartic acid [14] and cytosine [15] from cyanoacetylene. Additionally, the reactions of nucleobases with the Michael acceptor acrolein have also been explicitly investigated [16].

4.2 Experimental Procedures

4.2.1 Materials

Barbituric acid, cyanuric acid, 2,4,6-triaminopyrimidine (TAP), and cyanuric acid were purchased from Acros Organics and used as received. Adenine, uracil, and 5-aminouracil were purchased from Millipore Sigma and used as received. 2,4,5,6-Tetraaminoyrimidine (TetAP) sulfate was purchased from Millipore Sigma. Prior to use, TetAP sulfate was recrystallized from 2 M NaOH under an inert atmosphere to give TetAP free base. Ribose, DL-glyceraldehyde, succinic anhydride, *N*-methylsuccinimide, ethyl acetate, γ-butyrolactone, ethyl thioacetate, γ-butyrothiolactone, and methyl vinyl ketone were purchased from Millipore Sigma and used as received. Glycolide was purchased from Millipore Sigma. Prior to use, glycolide was recrystallized from ethyl acetate.

4.2.2 Reactions of Nucleobases with Electrophiles

The reactions of nucleobases with electrophiles were all carried out in sealed glass vials with stir bars at 0.5 M in each reactant in water on a 0.5 mL scale in an oil bath at 65°C for 24 hours. The reactions of TAP and TetAP were set up by first preparing 0.5 M aqueous stock solutions of the nucleobase, distributing the stock in 0.5 mL aliquots into vials, and then adding 0.25 μmol of electrophile to the vial. The reactions of barbituric acid and cyanuric acid were set up by first preparing 0.5 M aqueous stock solutions of the nucleobase with 1 eq. of triethylamine, distributing the stock in 0.5 mL aliquots into vials, and then adding 0.25 μmol of electrophile to the vial. The reactions of adenine, uracil, melamine, and 5-aminouracil were prepared by adding 0.25 μmol of nucleobase to the vial, adding 0.5 mL of water, and then adding 0.25 μmol of electrophile to the vial. Once the

electrophile was added, the vial was immediately sealed and immersed in an oil bath. After the reactions were completed, aliquots were taken for UV-LC/MS analysis. If the reactions presented as a suspension after 24 hours, the mixtures were vortexed rapidly before sampling the suspension. Aliquots were diluted 1:50 for UV-LC/MS analysis.

4.2.3 Analytical Techniques

After dilution, 2 μL samples were injected onto a 3.5 μm XBridge amide column running a linear gradient of 90% MeCN/10% NH_4OAc buffer 10 mM pH 9 to 60% MeCN/40% NH_4OAc buffer 10 mM pH 9 over a period of 7 minutes with a flow rate of 0.5 mL/min.

4.3 Experimental Selection of Model Nucleobases and Model Electrophiles

4.3.1 Selection of Model Nucleobases

The hexad-forming noncanonical nucleobases (TAP, melamine, barbituric acid, and cyanuric acid, Figure 1.4) were previously considered strong candidates for proto-nucleic acid recognition units based on their propensities to react with ribose and other sugars and on their propensities for self-assembly [17-19]. For these reasons, they are included in the set of nucleobases to be tested. These noncanonical nucleobases can be partitioned into two distinct families based on their hydrogen-bonding patterns: the donor-acceptor-donor family, (or the adenine-like family), and the acceptor-donor-acceptor family (or the uracil-like family). Because these nucleobases are expected to be forward-compatible with adenine and uracil, these canonical nucleobases themselves are also included.

Another notable class of nucleobases are the 5-aminopyrimidines. Amino groups at position 5 of a pyrimidine are more nucleophilic than amino groups at positions 2, 4 or 6 [20], as the latter amino groups are somewhat depleted in electron density by the withdrawing effect of the endocyclic nitrogen atoms by resonance. This is not the case for amino groups at position 5. Furthermore, substitution of 5-aminopyrimidines with electron-donating substituents at positions 2, 4, or 6 greatly enhance the nucleophilicity of position 5. This can be rationalized with a molecular orbital argument.

As an example, consider an isolated fragment of a 4,5-diaminopyrimidine molecule that consists of carbons 4 and 5 and their substituent amino groups (i.e., a 1,2-diaminoethylene fragment). Among these four atoms, there are six pi electrons. By constructing molecular orbitals from a basis set of the p orbitals that are perpendicular to the molecular plane of these four atoms, we find that the highest occupied molecular orbital (HOMO) of this system is antibonding (Figure 4.1). The result is the same when a 2,5-diaminopyrimidine system is considered. For a 1,2-disubstituted ethylene system (where the substituents are electron-donating), this phenomenon is described as the *vinylogous alpha effect*.

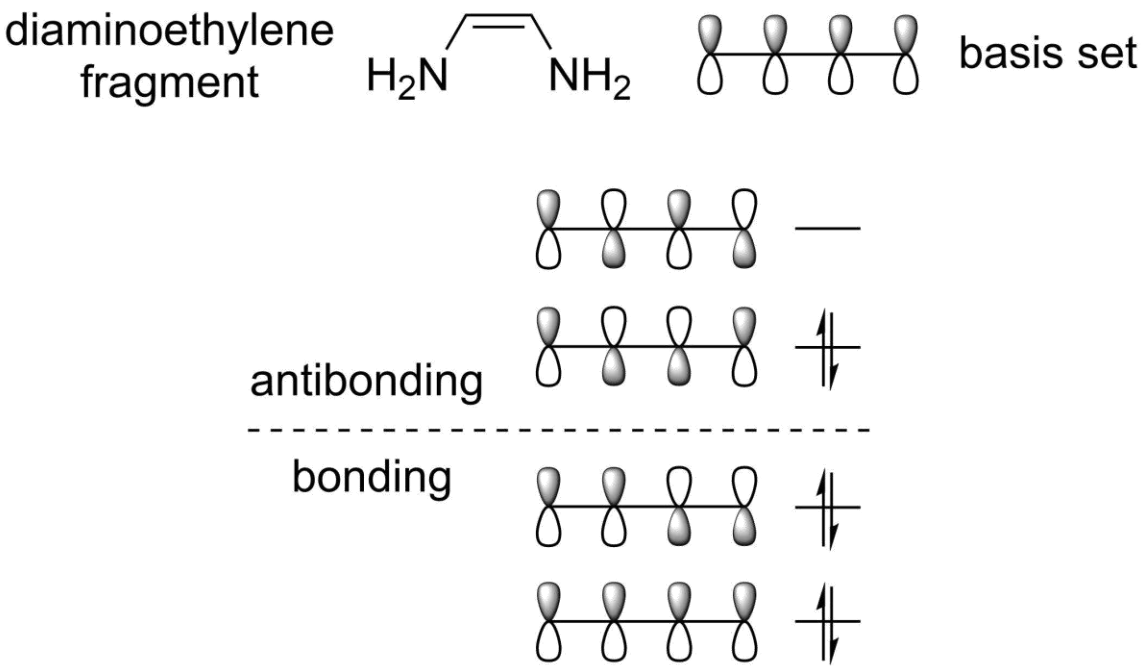


Figure 4.1. A demonstration of the vinylogous alpha effect for the 1,2-diaminoethylene fragment of a 4,5-diaminopyrimidine.

Because members of this class of pyrimidines are expected to be especially reactive, two specific examples are included in this survey: 2,4,5,6-tetraminopyrimidine (TetAP) and 5-aminouracil (5AU). TetAP was chosen for its obvious analogy to TAP. The corresponding analogue of barbituric acid, uramil (5-aminobarbituric acid), was not selected due to its profound insolubility in water. 5AU is sparingly soluble, but would perhaps be driven into solution as it is consumed.

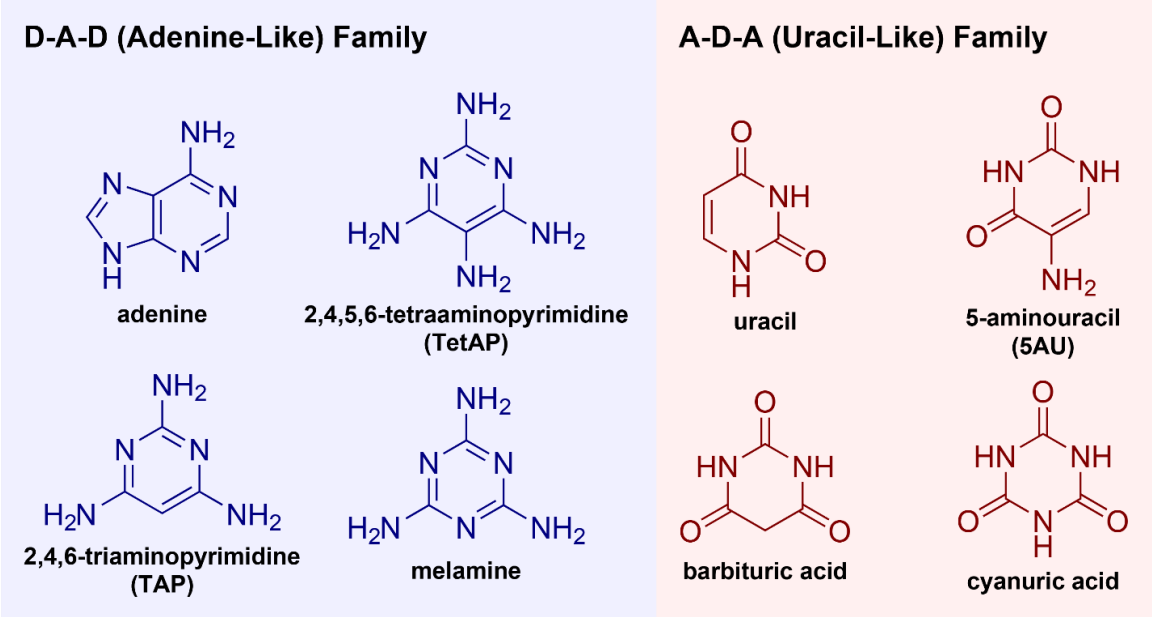


Figure 4.2. The set of nucleobases assessed in this study sorted according to hydrogen-bonding pattern. Left: the donor-acceptor-donor (D-A-D), or adenine-like family. Right: the acceptor-donor-acceptor (A-D-A), or uracil-like family.

4.3.2 Selection of Model Electrophiles

As was explained above, the aldehyde, anhydride, ester, imide, thioester, and certain Michael acceptors are all considered plausibly prebiotic electrophiles. However, different compounds that contain the same functional groups may display different patterns of reactivity. Therefore, for some electrophilic functional groups, more than one compound was tested.

Aldehydes, for example, are implicated in the reactivity of aldose sugars [21, 22]. However, for aldoses of four carbon atoms and greater, cyclic hemiacetal forms dominate over the free aldehyde [21]. For this reason, two test aldehydes were chosen: glyceraldehyde, a triose, which cannot adopt a cyclic form, and ribose, a pentose, which adopts four cyclic forms that dominate over the free aldehyde form.

Similarly, the reactivity of esters and thioesters may vary depending on the exact nature of the molecule containing the moiety. For esters, three compounds were chosen: ethyl acetate, a prototypical ester, γ-butyrolactone (GBL), a prototypical lactone, and glycolide (2,5-dioxo-1,4-dioxane), the cyclic dimer of glycolic acid, the simplest hydroxy acid. Among all possible lactones, GBL and glycolide were chosen for specific reasons. Unlike all other simple, isolable, n-membered lactones (e.g., n = 4, β-propiolactone; n = 6, δ-valerolactone; and n = 7, ε-caprolactone) GBL does not easily polymerize at room temperature [23]. This property is attributed to its low ring strain compared to the other simple lactones. Conversely, glycolide readily polymerizes upon heating or in the presence of a catalyst such as stannous octoate or 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) [24]. Furthermore, as the cyclic dimer of glycolic acid, glycolide itself may have prebiotic relevance. It is also analogous in structure and reactivity to morpholine-2,5-dione (the parent compound and derivatives thereof), the cyclic heterodimer of glycolic acid and glycine [25], which is also prebiotically relevant in the context of depsipeptide formation [26]. The differences in propensity for polymerization of GBL and glycolide may correlate to their electrophilicities.

For thioesters, two compounds were chosen: ethyl thioacetate, the thioester analogue of ethyl acetate, and γ-butyrothiolactone (GBTL), the thiolactone analogue of GBL. The hypothetical thiolactone of glycolide, 2,5-dioxo-1,4-dithiane (the cyclic dimer of thioglycolic acid), as it is not expected to form in a plausibly prebiotic manner. The monomer, thioglycolic acid, and derivatives thereof, are not expected to form by the same mechanism as α-amino acids and α-hydroxy acids, as this would require the presence of a thioaldehyde, which does not readily form due to the instability of the S-C double bond

(with no stabilizing electron-donating substituents). Reflecting its difficulty of formation, it is also not commercially available.

The formation of anhydrides and imides in a prebiotic manner is most plausible when occurring intramolecularly. For example, depsipeptide oligomers that contain either aspartic acid or malic acid residues may form intramolecular (5-membered ring) anhydrides if they are present at the C-terminus of the oligomer. Similarly, if an aspartic acid or malic acid residue forms a peptide bond with an amino acid in the *C*-terminal direction, intramolecular (5-membered ring) imides can be formed [27, 28]. For this reason, succinic anhydride and *N*-methylsuccinimide (NMS) were chosen as models for these types of moieties. If a nucleobase can be appended to either of these moieties, this would imply that the formation of proto-nucleic acid backbones could occur by depsipeptide oligomerization, and that the complete proto-nucleic acid could be subsequently formed by attachment of sufficiently reactive nucleobases to this pre-formed backbone.

Finally, for Michael acceptors, methyl vinyl ketone (MVK) was chosen, as the ketone moiety of this compound is not expected to be excessively electrophilic so that the primary mode of reactivity will be through conjugate addition (i.e., 1,4-addition).

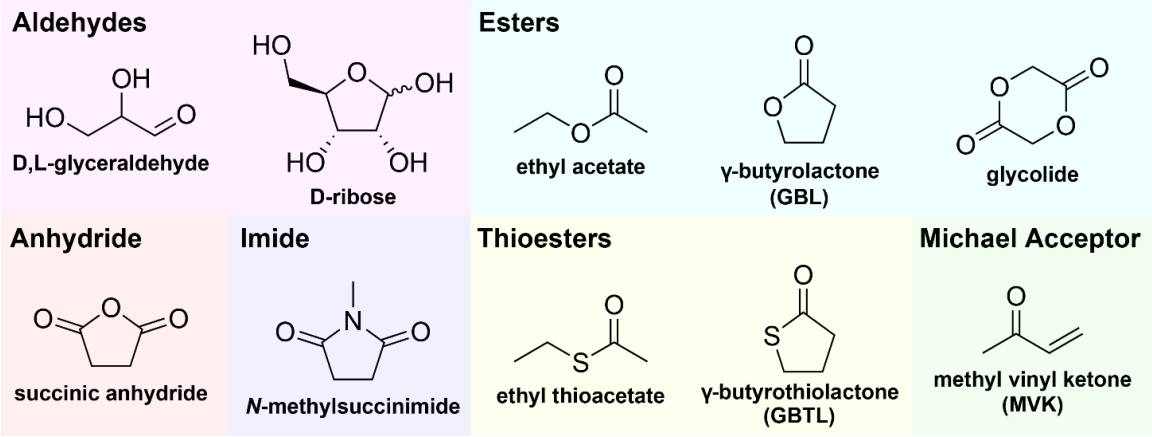


Figure 4.3. The set of electrophiles assessed in this study. Representatives from six classes of electrophiles (aldehydes, esters, anhydrides, imides, thioesters, and Michael acceptors) were chosen.

4.4 The Reactions of Nucleobases with Prebiotic Electrophiles

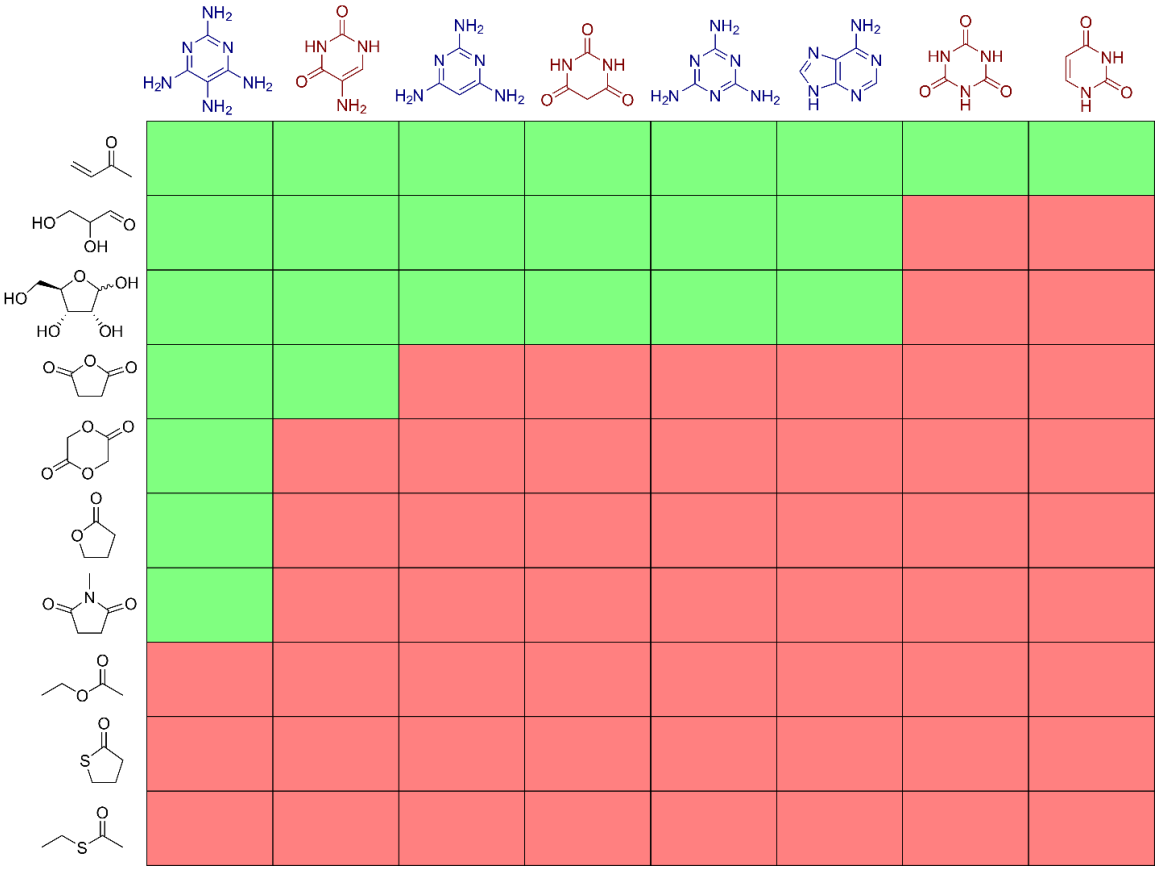
4.4.1 Results

With eight nucleophilic nucleobases and ten electrophiles, 80 total reactions were performed. The reactions were standardized by conducting all in water at 65°C at 0.5 M in each reactant for 24 hours. For members of the adenine family, no pH adjustment was made. As described above, members of the uracil family are expected to be most nucleophilic when deprotonated. Furthermore, it was found that barbituric acid and cyanuric acid were freely soluble in water when 1 equivalent of triethylamine was added. This solubilized effect was emphatically not the case when 1 equivalent of sodium or potassium hydroxide were added. Although these strong bases increased the solubility of barbituric acid and cyanuric acid, they did not solubilize them up to 0.5 M. Therefore, for all reactions involving members of the uracil family, 1 equivalent of triethylamine was added. This did not fully solubilize uracil and 5AU, so these reactions were run as

suspensions. Similarly, adenine and melamine were not soluble at 0.5 M, and were also run as suspensions.

The results of this set of reactions is summarized in Table 4.1. Green boxes indicate a reaction that produced at least one detectable product from a UV absorption peak by UV-LC/MS (except for cyanuric acid, which has no appreciable absorbance in the near-UV, and for which product formation was detected by mass spectrometry). Red boxes indicate reactions where no product was detected.

Table 4.1. Propensities of model prebiotic nucleobases to react with model prebiotic electrophiles. Green boxes indicate a successful reaction: at least one product is detected from a new UV absorption peak by UV-LC/MS (except for cyanuric acid, for which products were detected by a new peak in the total ion chromatogram). Red boxes indicate no detectable product formation.



MVK is the most potent electrophile, reacting with all nucleobases tested. MVK is a known mutagen, reacting with the nucleobase moiety of nucleotides in DNA, primarily through conjugate addition, but also through imine formation. For example, in guanine deoxyribonucleotides, MVK forms two types of adducts: a cyclic product formed from conjugate addition at position N1 of guanine, followed by imine formation with the 2-amino group, and a linear product formed by conjugate addition at position N7 [29]. Similar modes of reactivity are possible for the nucleobases surveyed here. Conjugate addition may proceed by nucleophilic attack by the endocyclic nitrogen atoms of any nucleobase, or by the exocyclic amino groups of members of the adenine family or 5AU. Imine formation is also possible at any exocyclic amino group of members of the adenine family or 5AU.

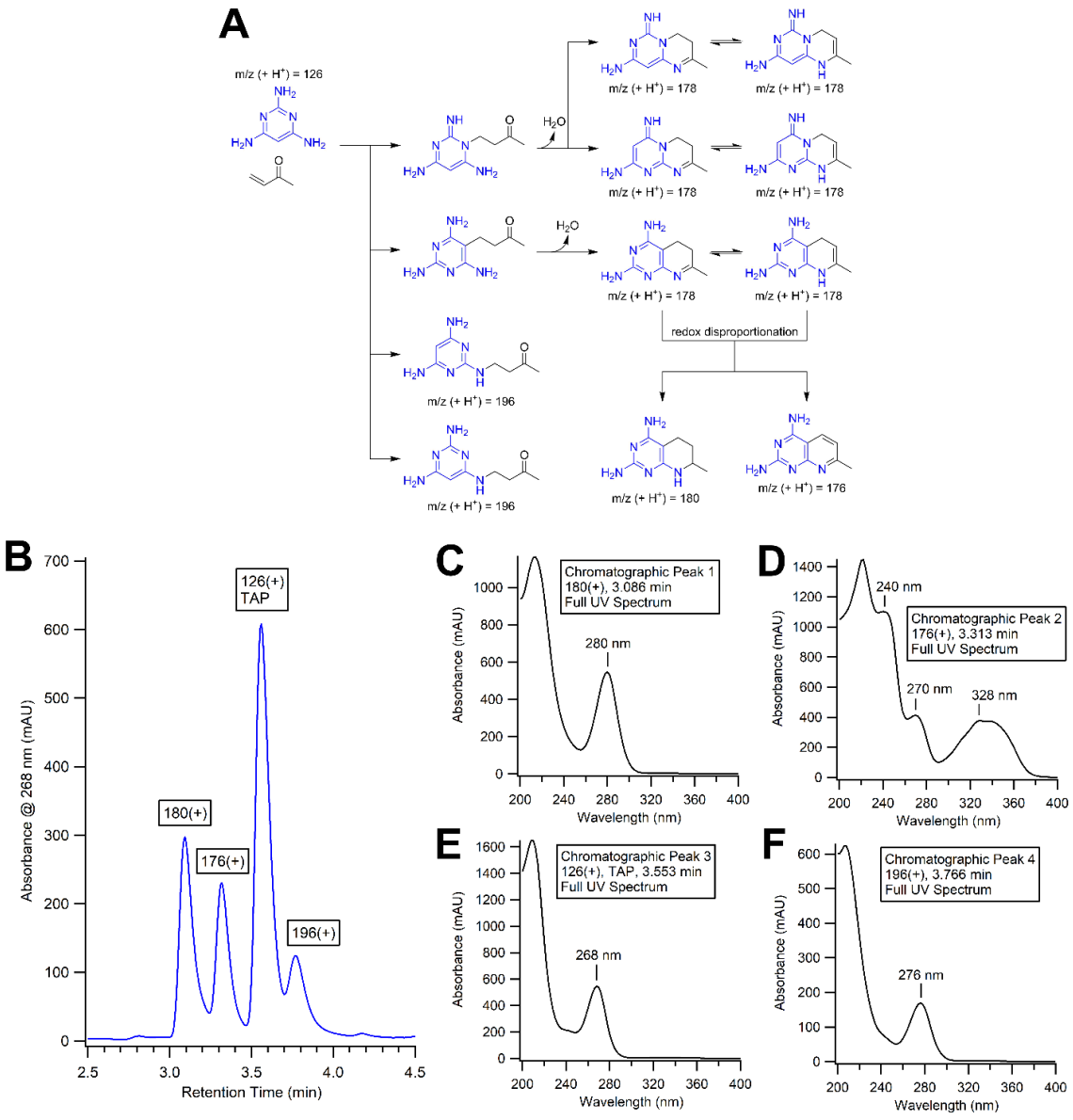


Figure 4.4. Analysis of the reaction of TAP with MVK. A. A proposed map explaining the products of the reaction of TAP with MVK. B. UV-LC/MS chromatogram of the reaction of TAP with MVK. Peaks are annotated with the corresponding masses detected in positive ion mode. C. UV Spectrum of chromatographic peak 1. D. UV Spectrum of chromatographic peak 2. E. UV Spectrum of chromatographic peak 3. F. UV Spectrum of chromatographic peak 4.

Reflecting this versatility in reactivity, all reactions with MVK gave multiple products. Exemplifying this, the reaction of TAP with MVK produces at least three products. A proposed reaction map and chromatographic data are shown in Figure 4.4. The latest-eluting peak in the UV-LC/MS chromatogram displays a mass of 196 ($m_{\mathrm{TAP}} +$

$m_{\mathrm{MVK}} + m_{\mathrm{H^+}} = 125 + 70 + 1 = 196$), consistent with a conjugate addition adduct of TAP with MVK. A number of sites on TAP can be substituted to give a product with this mass. The peak eluting second represents a product with a mass 20 units lower than the conjugate addition product (176). This mass may correspond to a conjugate addition product that has cyclized by condensation, followed by oxidation, perhaps by O_2 in air or by disproportionation (see below). Consistent with this hypothetical expansion of the aromatic system, the UV absorption spectrum shows a new, broad band with a local absorption maximum at 328. Neither TAP nor the other products of this reaction show this band. The first peak to elute gives an anomalous mass of 180 (in positive ion mode), 2 units higher than what would be expected for the product formed by conjugate addition and condensation. This mass may result from reduction, for example, of the imine moiety, possibly by disproportionation (also giving the oxidized product mentioned above). The reactions of MVK with the other nucleobases show similar trends: conjugate additions products (single, double, or triple additions) and condensation products are observed.

Glyceraldehyde and ribose show similar trends in reactivity. Given the extensive work that has been previously described (see § 2.3.2) on the addition of noncanonical nucleobases to ribose, the results found here are as expected. As cyanuric acid and uracil are incapable of forming a double-bonded condensation intermediate with ribose, they are not ribosylated. Adenine is known to be ribosylated in water at the exocyclic amino group at position 6 [30]. This is consistent with the previous analysis, as adenine is capable of forming a double-bonded intermediate with ribose through this group and only this group.

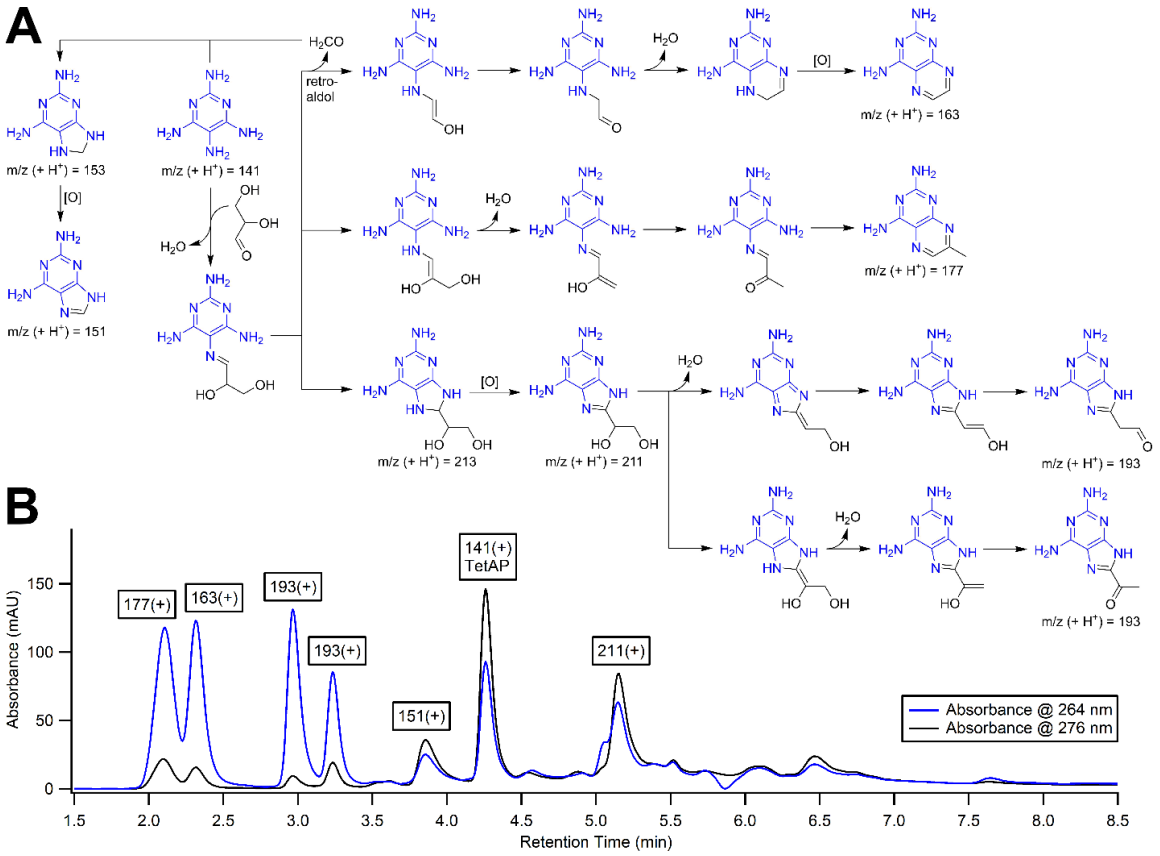


Figure 4.5. Analysis of the reaction of TetAP with glyceraldehyde. A. A proposed map explaining the products of the reaction of TetAP with glyceraldehyde. B. UV-LC/MS chromatogram of the reaction of TetAP with glyceraldehyde. Peaks are annotated with the corresponding masses detected in positive ion mode.

The success or failure of the reactions of the nucleobases with glyceraldehyde is consistent with the proposed mechanism of nucleobase glycosylation: only those products that can condense with the aldehyde moiety (i.e., react with the release of water) to form a double-bonded product or intermediate successfully react with glyceraldehyde. However, the fate of these products or intermediates is not known. Of the nucleobases that successfully reacted with glyceraldehyde, only barbituric acid cannot react by imine formation; rather, a Knoevenagel condensation gives the immediate product. For either an imine or Knoevenagel product, isomerization is possible. Furthermore, aldol and retro-aldol reactions may occur that further complicate the outcome. Take, for example, the

reaction of glyceraldehyde with TetAP. Many products are formed from this reaction (a map of which is shown in Figure 4.5 with a corresponding annotated UV-LC/MS chromatogram), some of which may result in ring closures that expand the aromatic system. The immediate product of the reaction, the condensation product (m/z = 213 in positive ion mode, for which several isomers are possible), is not observed. However, a mass at 211 is observed that corresponds to the oxidation of this product to a 2,6-diaminopurine derivative ($m_{\mathrm{TetAP}} + m_{\mathrm{glyceraldehyde}} - m_{\mathrm{H_2O}} - m_{\mathrm{H_2}} + m_{\mathrm{H^+}} = 140 + 90 - 18 - 2 + 1 = 211$). (Recall from Figure 4.1 that the HOMO of TetAP (and its derivatives) is antibonding, and is therefore high in energy and susceptible to oxidation by O_2 and other oxidizing agents.) At least two products are formed that result from condensation, oxidation, and further water elimination. One product is formed which corresponds to the mass of 2,4-diaminopteridine (m/z = 163). This product may result from an initial condensation, probably with the amino group at position 5 of TetAP, followed by retro-aldol eliminate to release formaldehyde, with subsequent tautomerization to give the Amadori product, which cyclizes by condensation with an adjacent amino group at position 4/6, and is finally oxidized to give the fully aromatic bicyclic product. A product with the mass of methyl-2,4-diaminopteridine (m/z = 177) is seen eluting as well. This may result from a redox-neutral reaction in which glyceraldehyde, or the imine condensation product with TetAP, undergoes β-elimination to form methylglyoxal (or its corresponding imine), which cyclizes by condensation with TetAP to form the 7-methyl-2,4-diaminopteridine product. Finally, formaldehyde released from retro-aldol reactions can react with TetAP to give a cyclic product, which is oxidized to give 2,6-diaminopurine (m/z = 151). Consistent with

this analysis, the differences is ratios of absorbances at 264 nm and 276 nm indicate that different heterocyclic aromatic cores are being produced.

The reaction of succinic anhydride with the 5-aminopyrimidines is successful in water, perhaps because these nucleobases (and only these nucleobases) are sufficiently strong nucleophiles to out-complete water for attack of the anhydride. Only monosubstituted products are observed, consistent with the hypothesis that the amino groups at position 5 of 5AU and TetAP are more nucleophilic than any other positions.

TetAP, as the most potent nucleophile, successfully reacts with lactones, and even NMS, to form monosubstituted products by ester aminolysis or transamidation, respectively. It is interesting to note that the reactions of TetAP with GBL and NMS were successful where the reaction of TetAP with GBTL (or with ethyl thioacetate), ostensibly the stronger electrophile, failed. Similarly, the absence of products from the attempted reaction of TetAP with ethyl acetate is interesting given that TAP successfully reacted with GBL. The lack of reactivity observed for these electrophiles may simply be due to insolubility, as ethyl acetate, ethyl thioacetate, and GBTL are all liquids that are immiscible with water.

4.5 Discussion

A great deal of insight into the spontaneous prebiotic emergence of informational polymers can be gleaned from this survey; however, a number of caveats must be stated.

First, the exact chemical structures of the products of these reactions have not yet been determined, and have only been rationalized according to physical organic principles

and previous studies. Isolation of the reactions products and structural determination by, for example, NMR, will be performed in the future to confirm the hypotheses on structures stated above.

Second, the uniform reaction conditions employed in this study were chosen to make meaningful comparisons in reactivity, but are by no means the only possible prebiotic reaction conditions. For example, as will be described in Chapter 5, TetAP is capable of reacting with soluble thioesters in solution to form amides. Another set of prebiotically relevant reaction conditions not studied here are the so-called "dry-down" reactions. These reactions, performed by drying aqueous solutions of reactants to deposit semi-liquid films, are commonly invoked in prebiotic synthesis [11, 26, 31]. Dry-down conditions have the advantage of driving condensation-dehydration reactions forward thermodynamically by the evaporation of water, allowing bonds to be formed spontaneously that would not have otherwise in aqueous solution [31]. A corollary to this is that, in the dried, semi-liquid state, reactions can occur that involve nucleophiles that are similar in strength to, or weaker than, water, as water is not present to out-compete them.

Finally, the reactions studied here were all allowed to proceed at 65°C for 24 hours. It may be possible that products between the tested nucleophiles and electrophiles do initially form, but are then hydrolyzed. However, if this were the case, the prebiotic significance of these reactions for the formation of proto-nucleic acids is dubious, because the products formed would not be stable enough to sustain an informational system in water.

The propensity for a reactant to participate in a reaction is often described by the "strength" of its nucleophilic or electrophilic moieties. This jargon is intuitively useful to organic chemists; an anhydride is typically thought of as a "stronger" electrophile than an ester, meaning that the anhydride is more likely than the ester to successfully react with a nucleophile to give a product, or to give a product in a higher yield. However, the "strength" of a nucleophile or an electrophile is not a singular parameter in describing reactivity. A familiar way to demonstrate this is through the recollection of the principle of hardness/softness [32]. "Hard" nucleophiles and electrophiles react with each other primarily through electrostatic attraction, while "soft" nucleophiles react with each other primarily due to a greater polarizability or orbital overlap. Although some soft nucleophiles are considered "strong", they may not participate in certain reactions that "weaker" hard nucleophiles participate in. For example, Fischer esterification is commonly employed for the acylation of alcohols, but is rare for thiol acylation; rather, anhydrides are typically required to acylate thiols [33], even though mercapto groups are typically thought of as "stronger" nucleophiles than hydroxyl groups.

The hardness/softness parameter is primarily concerned with the initial attack of an electrophile by a nucleophile, and is not concerned with the further details of the mechanism. Another criterion that takes into account the full mechanism of a reaction, and that is especially important for the reactions of nitrogenous heterocyclic compounds, is the *valency* of a nucleophile (or electrophile). This principle has been alluded to in this text in the explanation of the ability of certain noncanonical nucleobases to react with ribose and glyceraldehyde and the inability of the canonical nucleobases to do so. The valency of a nucleophile is the greatest number of bonds that it can possibly form with one or more

electrophiles in some intermediate or product in a reaction mechanism (and similarly for an electrophile with one or more nucleophiles). Although this definition is simple and obvious, it has great power in explaining the outcome of a reaction.

Consider, for example, the successful reaction of barbituric acid with ribose (Figure 2.5D), and the unsuccessful reaction of cyanuric acid with ribose (Figure 2.5E). The reaction of barbituric acid with ribose to form nucleosides is an example of a Knoevenagel condensation followed by intramolecular conjugate addition. Even though the product of the reaction is a riboside in which barbituric acid is bound to ribose by a single bond, the mechanism of this reaction requires the formation of a double-bonded intermediate. Barbituric acid, as a *divalent* nucleophile, is capable of forming this intermediate. The divalency of barbituric acid is obvious, as the parent compound is substituted with two hydrogen atoms at position 5 (the nucleophilic position). At first, it might seem that cyanuric acid can also act as a divalent nucleophile, as the nucleophilic site, the imidic nitrogen atom, formally possesses to electronic groups: a hydrogen substituent, and a lone pair of electrons. However, the lone pair of the imidic nitrogen atom is ineffective as a nucleophile, as it is part of the conjugated π system. Only upon deprotonation can this site act as a (monovalent) nucleophile, and emphatically by the lone pair revealed by deprotonation that is parallel to the plane of the aromatic ring. (It is important to note that, even though cyanuric acid and the barbiturate anion are isoelectronic, the lone pair at position 5 of the barbiturate anion retains its nucleophilicity, as the lower electronegativity of the carbon atom of the pyrimidine ring destabilizes this lone pair relative to the lone pair of the nitrogen atom of the triazine ring of cyanuric acid.) Therefore, although cyanuric acid can, in principle, attack the aldehyde moiety of ribose, it cannot further advance, as it

is incapable of expelling water to form a double-bonded intermediate. This same principle applies to the canonical nucleobases and their failure to react with ribose to form canonical nucleosides, as the sites of canonical glycosidic bond formation are monovalent nucleophiles. It should be noted that, if the mechanism of glycosylation proceeds by an oxocarbenium intermediate (see Figure 3.7), which may be the case with harsh reaction conditions, only a monovalent nucleophile is required.

This principle of valency can also be applied to the present survey of nucleophilic nucleobases and prebiotic electrophiles. Adenine, melamine, barbituric acid, TAP, 5AU, and TetAP all contain divalent nucleophilic sites. These are either exocyclic amino groups or, in the case of TAP and barbituric acid, endocyclic carbon nucleophiles at position 5. Adenine also contains four monovalent nucleophilic sites: N1, N3, N7, and N9. Because N9, the site of canonical nucleosidation, is monovalent, the ribosides formed from the reaction of adenine and ribose in water do not contain adenosine. Rather, these ribosides are formed from condensation with the amino group at position 6 [30]. All of the pyrimidines contain monovalent nucleophilic sites at N1 and N3. Similarly, the triazines contain monovalent nucleophilic sites at N1, N3 and N5. Besides 5AU, which contains a divalent nucleophilic exocyclic amine, all of the nucleophilic sites on members of the uracil family require deprotonation to activate them as nucleophiles.

As a Michael acceptor, MVK acts as a monovalent electrophile. Reflecting this, all nucleobases studied form adducts with MVK. MVK can also act as a divalent electrophile at its ketone moiety. As expected, dehydration products with MVK can be seen for all members of the adenine family, which probably result from initial conjugate addition, followed by condensation cyclization (see Figure 4.4 for an example of this with TAP). As

discussed in § 2.3.2 and § 4.4.1, the formation of stable products with glyceraldehyde or ribose requires a divalent nucleophile. Reflecting this, only cyanuric acid and uracil fail to react with these sugars.

The remainder of the electrophiles examined are all carboxylic acid derivatives. Put another way, the electrophilic sites of these compounds all feature carbonyl carbon atoms in the +3 oxidation state (as opposed to ketones (+2) and aldehydes (+1)). Although the mechanism of substitution of carboxylic acid derivatives only requires a monovalent nucleophile, the product formed derives stability from delocalization of lone pairs of electrons on the new substituent. Therefore, even if products can initially be formed between certain nucleophile-electrophile pairs of this class, they may be liable to hydrolysis if they are formed with monovalent nucleophiles. Divalent nucleophiles, on the other hand, can potentially form stable compounds by delocalization of electrons from their remaining electron group. However, the carboxylic acid derivatives studied here vary in their strengths as electrophiles. Exemplifying this, glycolide, GBL, and NMS (perhaps the poorest electrophile) only reacts with TetAP, the strongest nucleophile. Similarly, only 5AU and TetAP, which feature highly nucleophilic amino groups at position 5, form stable products with succinic anhydride. Even though TAP, melamine, barbituric acid, and adenine also feature divalent nucleophilic sites, these sites are depleted in electron density by the electron-withdrawing effect of nearby substituents, diminishing their absolute nucleophilic strength, or reducing the hydrolytic stability of the acyl product.

There are a number of important consequences for the prebiotic emergence of informational polymers. For all of the nucleobases studied, there is at least one, and often several, types of electrophile that it can successfully react with. The choice of candidate

proto-nucleic acid backbone, therefore, depends on the choice of candidate proto-nucleic acid nucleobase, and vice versa. For example, for a sugar-based proto-nucleic acid, only nucleobases containing divalent nucleophilic sites are viable, assuming that the glycosylation reaction proceeds by nucleophilic attack of the aldehyde of the open-chain form of the sugar, and that nucleoside is formed from the reaction of a pre-formed nucleobase with a pre-formed sugar.

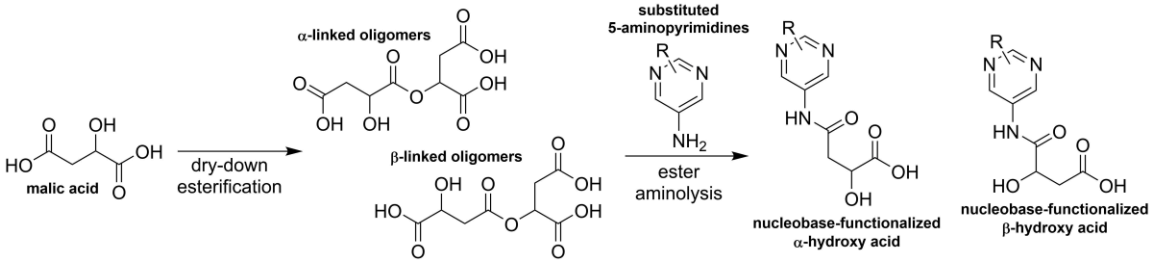


Figure 4.6. A proposed prebiotic route to nucleobase-functionalized α- and β-hydroxy acids. When dried from an acidic aqueous solution, malic acid spontaneously oligomerizes by esterification. Oligomers containing α linkages and β linkages are formed. These esters can be aminolyzed by 5-aminopyrimidines, either in the dry or rehydrated state, to give nucleobase-functionalized α- and β-hydroxy acids.

Given that TetAP successfully attacks esters, any proto-nucleic acid backbone that contains side-chain carboxylic acids that can be prebiotically esterified could potentially form a TetAP-functionalized polymer. Although 5AU, a potential pairing partner for TetAP, failed to react with esters under these conditions, it may be possible to react 5AU with esters under alternative, prebiotically relevant conditions. Esters, for example, are formed prebiotically by the dry-down reaction of α-hydroxy acids [11]. If TetAP and 5AU were present in solution with hydroxy acids with carboxylic acid-functionalized side-chains (such as malic acid), a dry-down process might produce 5AU- and TetAP-functionalized oligoesters. In this specific example, anhydrides may also form under these

drying conditions, which emphatically permits the attachment of both 5AU and TetAP (Figure 4.6). The viability of this system will be discussed in Chapter 5.

It is interesting to note that many types of nucleobase successfully react with Michael acceptors such as MVK. Cleaves and others have studied the reactions of nitrogenous heterocycles with another Michael acceptor, acrolein, and have found that it also robustly alkylates these compounds [16]. Recently, Rodriguez et al. found that when spark discharge experiments are performed above an aqueous solution of a nucleobase (53 of which were surveyed from 5 classes of nitrogenous heterocycle), the primary products are those formed from conjugate addition reactions with Michael acceptors such as acrolein, acrylonitrile, and cyanoacetylene [34]. Because the process of conjugate addition of nucleobases to Michael acceptors is facile and robust, the products of these reactions should perhaps receive stronger consideration when formulating hypotheses on the spontaneous prebiotic emergence of genetic polymers. This is especially apparent when the challenges associated with prebiotically forming canonical nucleotides are also considered.

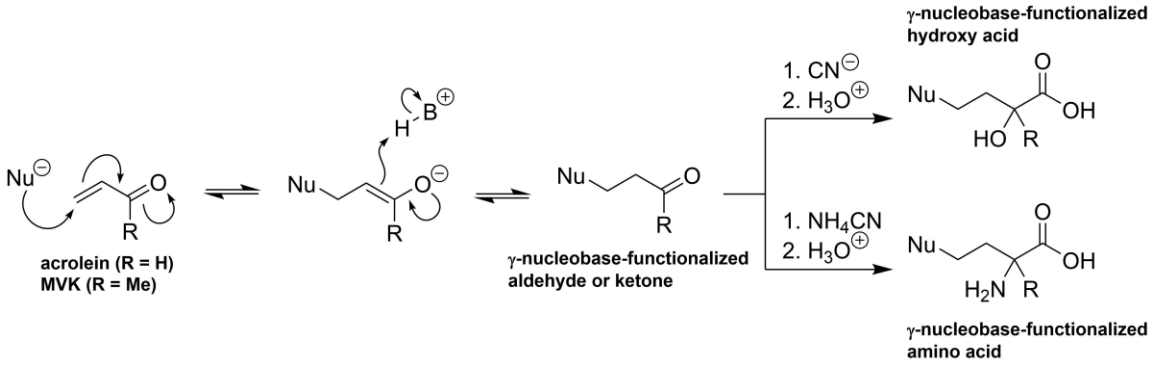


Figure 4.7. A proposed prebiotic route to nucleobase-functionalized α-amino acids and α-hydroxy acids. A nucleobase (Nu) attacks acrolein (R = H) or MVK (R = Me) to form a nucleobase-functionalized aldehyde or ketone. This carbonyl compound can form a hydroxy acid through the cyanohydrin pathway (attack of cyanide to form a

cyanohydrin, which is then hydrolyzed to form a hydroxy acid) or an amino acid through the Strecker synthesis (imine formation, followed by attack of cyanide to form an amino nitrile, which is then hydrolyzed to form an amino acid).

The immediate products formed from the conjugate addition of nucleobases to MVK or acrolein are γ-nucleobase-functionalized butan-2-ones and propionaldehydes, respectively, both of which are substrates for the Strecker synthesis of α-amino acids [35] and synthesis of α-hydroxy acids via cyanohydrin formation (see Chapter 6). With the appropriate cyanide source that does not out-compete the nucleobases for addition to acrolein or MVK, it may be possible for nucleobase-functionalized amino acids and hydroxy acids to be formed prebiotically (Figure 4.7). Because these two classes of compounds both contain a nucleophilic moiety (an amino group or a hydroxyl group) and an electrophilic moiety (a carboxylic acid), they have the potential to form nucleobase-functionalized oligomers. (Interestingly, the reactions of 5-aminopyrimidines with malic acid, and the conjugate addition of nucleobases to acrolein, followed by cyanohydrin formation and hydrolysis, both produce nucleobase-functionalized hydroxy acids.) The relevance of nucleobase-functionalized amino acids and hydroxy acids and their possible prebiotic oligomerization will be discussed in Chapters 5 and 6.

4.6 Conclusions

Historically, the formation of canonical nucleosides (or nucleotides) has been seen as the primary obstacle to the prebiotic synthesis of RNA [36]. In the search for prebiotically plausible mechanisms of canonical nucleoside formation, the enormous chemical space of associated with the formation of noncanonical nucleic acids has been largely ignored (with some exceptions [1, 37, 38]). In contrast to the direct ribosylation of the canonical nucleobases, which is prebiotically difficult, the reactions of some alternative

nucleobases with ribose, or different classes of electrophiles, is facile, even in water. This suggests that the spontaneous prebiotic emergence of alternative genetic polymers is more plausible than the spontaneous prebiotic emergence of RNA. The exact identity of the proto-nucleic acid remains unknown, but clues to its chemical nature can be extracted from the physical organic principles elucidated here. Further work will investigate proto-nucleic acid monomer formation from reactions similar to those investigated here, and oligomerization of these candidate monomers to form genetic polymers.

4.7 References

[1] Hud, N.V., et al., The origin of RNA and 'My Grandfather's Axe'. Chem. Biol., 2013. 20: p. 466-474.

[2] Rios, A.C. and Y. Tor, On the Origin of the Canonical Nucleobases: An Assessment of Selection Pressures across Chemical and Early Biological Evolution. Israel Journal of Chemistry, 2013. 53(6-7): p. 469-483.

[3] Rios, A.C., H.T. Yu, and Y. Tor, Hydrolytic fitness of N-glycosyl bonds: comparing the deglycosylation kinetics of modified, alternative, and native nucleosides. Journal of Physical Organic Chemistry, 2014: p. 173-180.

[4] Baumann, M., et al., Synthesis of riboflavines, quinoxalinones and benzodiazepines through chemoselective flow based hydrogenations. Molecules, 2014. 19(7): p. 9736-59.

[5] Menor-Salvan, C., et al., Synthesis of pyrimidines and triazines in ice: Implications for the prebiotic chemistry of nucleobases. Chemistry-a European Journal, 2009. 15(17): p. 4411-4418.

[6] Callahan, M.P., et al., Carbonaceous meteorites contain a wide range of extraterrestrial nucleobases. Proceedings of the National Academy of Sciences of the United States of America, 2011. 108(34): p. 13995-13998.

[7] Segura, A., et al., Biosignatures from Earth-like planets around M dwarfs. Astrobiology, 2005. 5(6): p. 706-25.

[8] Miller, S.L., The mechanism of synthesis of amino acids by electric discharges. Biochimica Et Biophysica Acta, 1957. 23(3): p. 480-489.

[9] Breslow, R., ON THE MECHANISM OF THE FORMOSE REACTION. Tetrahedron Letters, 1959(21): p. 22-26.

[10] Botta, O. and J.L. Bada, Extraterrestrial organic compounds in meteorites. Surv. Geophys., 2002. 23(5): p. 411-467.

[11] Mamajanov, I., et al., Ester formation and hydrolysis during wet-dry cycles: generation of far-from-equilibrium polymers in a model prebiotic reaction. Macromolecules, 2014. 47: p. 1334-1343.

[12] de Duve, C., Blueprint for a Cell : The Nature and Origin of Life. 1991: New York: Patterson.

[13] Van Trump, J.E. and S.L. Miller, Prebiotic synthesis of methionine. Science, 1972. 178(4063): p. 859-60.

[14] Sanchez, R.A., J.P. Ferris, and L.E. Orgel, Cyanoacetylene in prebiotic synthesis. Science, 1966. 154(3750): p. 784-5.

[15] Robertson, M.P. and S.L. Miller, An efficient prebiotic synthesis of cytosine and uracil. Nature, 1995. 375(6534): p. 772-4.

[16] Cleaves, H.J., The Reactions of Nitrogen Heterocycles with Acrolein: Scope and Prebiotic Significance. Astrobiology, 2002. 2(4): p. 403-415.

[17] Chen, M.C., et al., Spontaneous prebiotic formation of a β-ribofuranoside that self-assembles with a complementary heterocycle. Journal of the American Chemical Society, 2014. 136(15): p. 5640-5646.

[18] Cafferty, B.J. and N.V. Hud, Was a pyrimidine-pyrimidine base pair the ancestor of Watson-Crick base pairs? Insights from a systematic approach to the origin of RNA. Israel Journal of Chemistry, 2015. 55(8): p. 891-905.

[19] Cafferty, B.J., et al., Spontaneous formation and base pairing of plausible prebiotic nucleotides in water. Nature Communications, 2016. 7: p. 11328.

[20] Becker, S., et al., A high-yielding, strictly regioselective prebiotic purine nucleoside formation pathway. Science, 2016. 352(6287): p. 833-836.

[21] Dworkin, J.P. and S.L. Miller, A kinetic estimate of the free aldehyde content of aldoses. Carbohydr. Res., 2000. 39: p. 359-365.

[22] Horenstein, N.A., Mechanisms for nucleophilic aliphatic substitution at glycosides. Adv. Phys. Org. Chem., 2006. 41: p. 275-314.

[23] Hong, M. and E.Y. Chen, Completely recyclable biopolymers with linear and cyclic topologies via ring-opening polymerization of gamma-butyrolactone. Nat. Chem., 2016. 8(1): p. 42-9.

[24] Chujo, K., et al., Ring-opening polymerization of glycolide. Die Makromolekulare Chemie, 1967. 100(1): p. 262-266.

[25] Martinez-Palau, M., et al., Morpholine-2,5-dione. Acta Crystallogr C, 2006. 62(Pt 5): p. o262-4.

[26] Forsythe, J.G., et al., Ester-Mediated Amide Bond Formation Driven by Wet-Dry Cycles: A Possible Path to Polypeptides on the Prebiotic Earth. Angewandte Chemie-International Edition, 2015. 54(34): p. 9871-9875.

[27] Clarke, S., Propensity for spontaneous succinimide formation from aspartyl and asparaginyl residues in cellular proteins. Int. J. Pept. Protein Res., 1987. 30(6): p. 808-21.

[28] Stephenson, R.C. and S. Clarke, Succinimide formation from aspartyl and asparaginyl peptides as a model for the spontaneous degradation of proteins. J. Biol. Chem., 1989. 264(11): p. 6164-70.

[29] Eder, E., C. Hoffman, and C. Deininger, Identification and characterization of deoxyguanosine adducts of methyl vinyl ketone and ethyl vinyl ketone. Genotoxicity of the ketones in the SOS Chromotest. Chem. Res. Toxicol., 1991. 4(1): p. 50-7.

[30] Fuller, W.D., R.A. Sanchez, and L.E. Orgel, Studies in prebiotic synthesis. VI. synthesis of purine nucleosides. J. Mol. Biol., 1972. 67(1): p. 25-33.

[31] Ross, D.S. and D. Deamer, Dry/Wet Cycling and the Thermodynamics and Kinetics of Prebiotic Polymer Synthesis. Life (Basel), 2016. 6(3).

[32] Pearson, R.G., Hard and Soft Acids and Bases. Journal of the American Chemical Society, 1963. 85(22): p. 3533-3539.

[33] Kazemi, M. and L. Shiri, Thioesters synthesis: recent adventures in the esterification of thiols. Journal of Sulfur Chemistry, 2015. 36(6): p. 613-623.

[34] Rodriguez, L.E., et al., Nitrogen heterocycles form peptide nucleic acid precursors in complex prebiotic mixtures. Sci. Rep., 2019. 9(1): p. 9281.

[35] Cleaves, H.J. and J.L. Bada, The Prebiotic Chemistry of Alternative Nucleic Acids, in Genesis - In The Beginning: Precursors of Life, Chemical Models and Early Biological Evolution, J. Seckbach, Editor. 2012, Springer Science+Business Media Dordrecht. p. 3-33.

[36] Orgel, L.E., Prebiotic chemistry and the origin of the RNA world. Critical Reviews in Biochemistry and Molecular Biology, 2004. 39(2): p. 99-123.

[37] Eschenmoser, A., Searching for nucleic acid alternatives. Chimia, 2005. 59(11): p. 836-850.

[38] Nelson, K.E., M. Levy, and S.L. Miller, Peptide nucleic acids rather than RNA may have been the first genetic molecule. Proceedings of the National Academy of Sciences of the United States of America, 2000. 97(8): p. 3868-3871.

CHAPTER 5. INVESTIGATIONS INTO PLAUSIBLY PREBIOTIC PROTO-NUCLEIC ACID BACKBONES

5.1 Introduction

The prebiotic synthesis of RNA is complicated by the difficulty in forming the ribose component of the backbone selectively [1] (as model prebiotic reactions that produce sugars, such as the formose reaction, give many products [2]) and the difficulty in esterifying phosphate to form a phosphodiester polymer [3] (as phosphoesterification is kinetically and thermodynamically unfavorable in water). Chemical scenarios have been proposed that increase the selectivity of the formose reaction for ribose that involve the addition of borate [4, 5], but this increase in selectivity is marginal. Similarly, plausibly prebiotic conditions have been discovered that facilitate phosphorylation of alcohols, such as ribose. However, these scenarios often involve hydrolytically unstable (and potentially prebiotically depleteable) activated forms of phosphate [6], or involve solvent conditions that would be deleterious to the formation of sugars and nucleosides, such as non-aqueous or semi-aqueous solvents that contain urea, ammonia, and/or formamide [7] (as sugars react readily with these nitrogen-containing compounds in a Maillard-like reaction to form tars [8]).

For these reasons, and others that suggest that RNA is an optimized structure, it may be that the extant nucleic acids are evolved, and were preceded by a series of pre-RNAs, all of which were ultimately descended from an ancestral proto-nucleic acid that formed more easily under prebiotic conditions [9]. As discussed in Chapter 1 (see § 1.3.2),

the chemical space of possible proto-nucleic acids is vast, but reasonable deductions can be made that reduce the size of this space. In Figure 1.5, a matrix of possible backbone linkages was considered from possible prebiotic nucleophiles and electrophiles. Some of these linkages can be considered less likely than others on the grounds of difficulty of prebiotic formation of hydrolytic instability. Others can be promoted or demoted in candidacy based on previous studies on related chemical linkages, or on the prebiotic synthesis of the relevant chemical moieties of a given linkage.

It is also important to note that, with any choice of candidate proto-nucleic acid backbone, geometric considerations must be made with respect to the mode of supramolecular association of the informational polymer. For example, if a backbone is examined that is appended with hexad-forming nucleobases, then this backbone must sterically permit the formation of the hexad to be functional.

In this chapter, initial efforts are described to assess the viability of candidate proto-nucleic acid backbones formed from two different types of chemical bonds: the thioaminal (i.e., the *N*,*S*-acetal) and the ester.

5.2 Experimental Methods

5.2.1 Materials

Cyanuric acid and 2,4,6-triaminopyrimidine (TAP) were purchased from Acros Organics and used as received. 2,4,5,6-Tetraaminoyrimidine (TetAP) sulfate was purchased from Millipore Sigma. Prior to use, TetAP sulfate was recrystallized from 2 M NaOH under an inert atmosphere to give TetAP free base. Ethanolamine, glyoxylic acid

monohydrate, pyruvic acid, DL-*N*-acetylhomocysteine thiolactone, 5-aminouracil (5AU), (R/S)-α-hydroxy-γ-butyrolactone (HBL), and 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) were purchased from Millipore Sigma and used as received.

5.2.2 Analytical Techniques

NMR spectra of thioaminal compounds were collected on a Bruker DRX 500 MHz NMR and were the sum of 64 transients. NMR spectra of Cy^{HA} were collected on a Bruker DRX 700 MHz NMR and were the sum of 64 transients (^{1}H) or 1024 transients (^{13}C).

UV-LC/MS analysis was performed with a 3.5 μm XBridge amide column running a linear gradient of 90% MeCN/10% NH_4OAc buffer 10 mM pH 9 to 60% MeCN/40% NH_4OAc buffer 10 mM pH 9 over a period of 7 minutes with a flow rate of 0.5 mL/min.

5.2.3 Synthesis

Thiazolidine-2-Carboxylic Acid (T2C) and 2-Methylthiazolidine-2-Carboxylic Acid (MT2C): These compounds were synthesized and purified according to a protocol described in reference [12].

1-(2,6-Diaminopurin-8-yl)-homocysteamine (DAPHC): Solid TetAP (free base, 140 mg, 1 mmol) and DL-*N*-acetylhomocysteine thiolactone (318 mg, 2 mmol) were mixed and pulverized into a fine, homogenous powder with a mortar and pestle. The powder was transferred to a glass vial with a magnetic stir bar. The vial was sealed and the contents were heated at 145°C with stirring for two hours. To the resulting deep red liquid, 1 mL of 2 M NaOH was added. The vial was sealed and the solution was heated at 155°C for two hours. The crude mixture was acidified with HCl and purified on a Teledyne Isco

CombiFlash Rf+ flash chromatography system using a Teledyne Isco C18Aq column with a 100% H_2O (0.1% TFA) eluent and a 100% MeCN column wash after product elution. Fractions containing only the product masses (as determined by ESI-MS) were lyophilized and combined to give DAPHC as the TFA salt.

2-Hydroxy-4-cyanuryl-butanoic acid (Cy^{HA}): To 100 mL of DMF, 6.45 g of cyanuric acid (50 mmol) and 7.48 mL of DBU (50 mmol) were added. The mixture was heated to 125°C with rapid stirring. Once the cyanuric acid had fully dissolved, 1.95 mL of DL-α-hydroxy-γ-butyrolactone (25 mmol) were added dropwise. The reaction was stirred at 125°C for 24 hours. The solution was then poured onto a watch glass and allowed to evaporate over several days, eventually producing a thick, brown oil above crystals of unreacted cyanuric acid. This material was suspended in 50 mL of water, and the suspension was centrifuged. The pellet was discarded and the supernatant was separated, in 5 mL portions, using a Teledyne Isco CombiFlash Rf+ purification system with a RediSep C18Aq 150g Gold column with a 100% water eluent. The fractions containing product, as detected by LC/MS, were pooled and reduced to about 10 mL with a rotary evaporator. To remove residual DBU, this solution was eluted over 25 g of Dowex 50WX8 cation exchange resin, hydrogen form, with 0.1% aqueous formic acid solution. 100 mL were collected and the solvent was removed with a rotary evaporator to give white flakes of racemic Cy^{HA} in its free acid form (1.73 g, 30%). ^{1}H NMR (700 MHz, D_2O): δ 4.33 (dd, 8.1 Hz, 4.1 Hz, 1H, H2), 3.96 (m, 2H, H4), 2.15 (m, 1H, H3), 2.01 (m, 1H, H3'). ^{13}C NMR (176 MHz, D_2O): δ 177.2 (C1), 150.9 (C5), 150.0 (C6), 68.1 (C2), 37.9 (C4), 31.0 (C3). HRMS (m/z): [M+H]+ $C_7H_{10}O_6N_3$, theoretical mass: 232.0564, actual mass: 232.0568.

5.3 Investigations into Thioaminal (*N*,*S*-Acetal) Backbone Linkages

Acetals have previously been investigated as a possible backbone linkage that preceded phosphodiesters [10, 11]. Computational studies indicate that the replacement of the phosphate moiety in RNA with a glyoxylate acetal (of either R or S orientation) does not significantly perturb the overall structure [11]. Thymidine dimers formed by a glyoxylic acid acetal can be formed under drying conditions with a divalent metal Lewis acid catalyst. Longer oligomers were not detected.

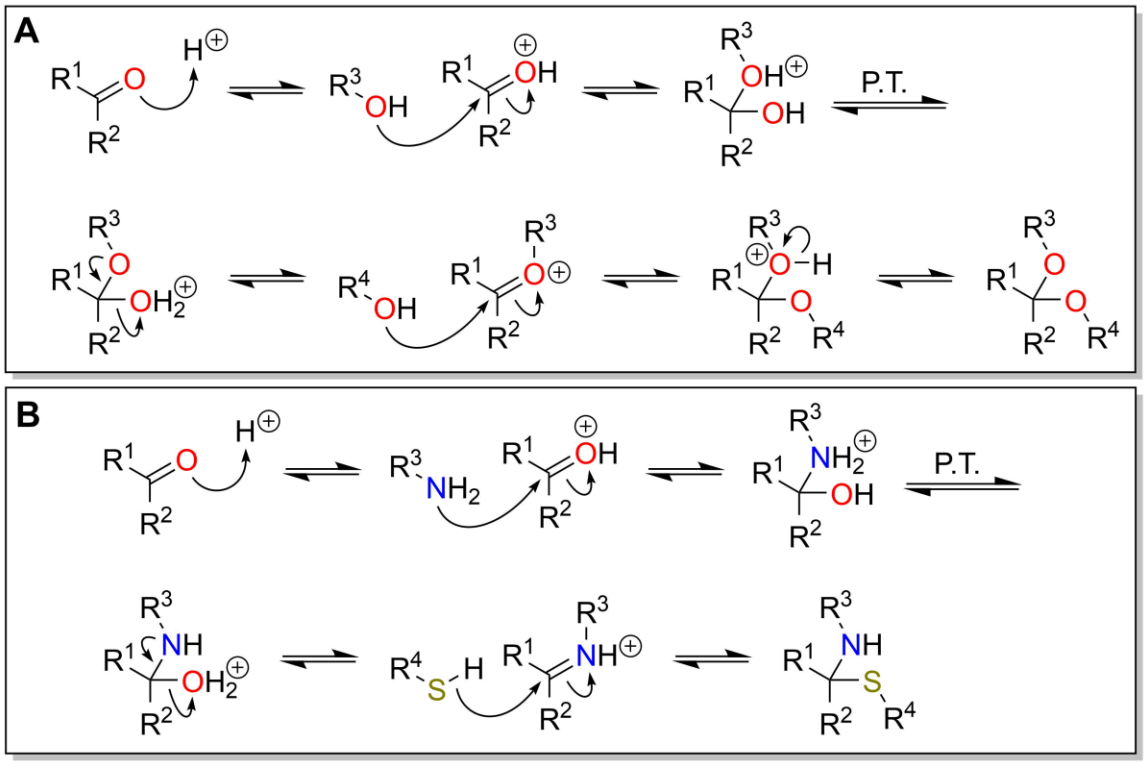


Figure 5.1. The mechanisms of formation of the acetal (A) and the thioaminal (B).

The acetal (or ketal) linkage is formed in an acid-catalyzed manner by an initial condensation of an aldehyde (or a ketone) with an alcohol to form an oxocarbenium intermediate which is then attacked by another alcohol (Figure 5.1A). The acetal, however, is not the only chemical moiety derived from an aldehyde that follows this same

mechanism. Rather than an alcohol attacking the aldehyde to form an oxocarbenium intermediate, an amine can attack to form an iminium intermediate. Due to the greater nucleophilicity of amines over alcohols, this iminium intermediate forms more readily than the oxocarbenium ion at higher pH values. This iminium species can be attacked by a number of nucleophilic species to form different types of bonds (see Figure 1.5, column 2). Alcohols compete poorly with water, complicating this reaction. Replacing an alcohol with the more nucleophilic thiol may allow the reaction to proceed in water. One subtle difference in the mechanism is that the thiol species probably requires prior or simultaneous deprotonation with nucleophilic attack, as thiolates are much stronger nucleophiles than neutral thiols. The linkage formed from this reaction is a *thioaminal*, or an *N*,*S*-acetal (Figure 5.1B). If a trifunctional connector of a proto-nucleic acid contained both amino and mercapto groups, it could potentially condense with an aldehyde such as glyoxylic acid (the conjugate base of which serves at the ionized linker) to form a thioaminal polymer.

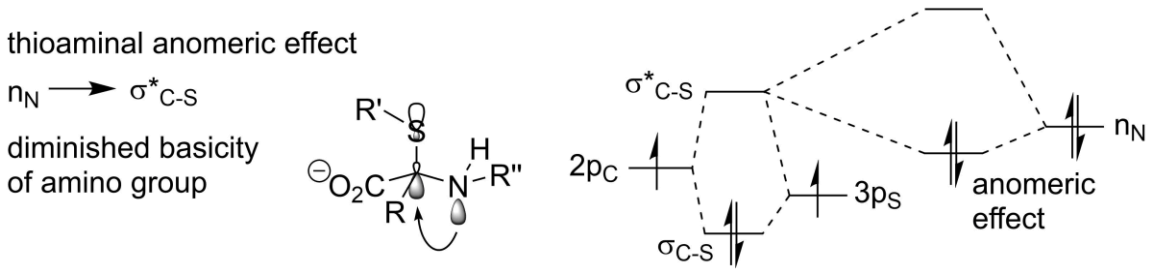


Figure 5.2. The anomeric effect for the thioaminal moiety. Because of the relatively low stabilization energy of the C-S σ bond, the C-S σ* orbital is not exceptionally high in energy, and orbital overlap can occur with the lone pair (n orbital) of the adjacent nitrogen atom, lowering its energy and reducing its basicity.

The thioaminal, like the acetal, was hypothesized to be a relatively stable chemical linkage due to electron delocalization by the anomeric effect, primarily in donation of electron density of the nitrogen lone pair to the C-S anti-bonding orbital (Figure 5.2). This

would also have the effect of diminishing the basicity of the nitrogen atom so that, at some pH range not far from neutrality, a negatively charged polymer (as opposed to a zwitterionic polymer) could be formed. To test the stability of the thioaminal linkage in water, several test compounds were synthesized and isolated, and their rates of formation and hydrolysis were measured by ^{1}H NMR.

Cyclic cysteamine thioaminals of glyoxylic acid (thiazolidine-2-carboxylic acid, T2C) and pyruvic acid (2-methylthiazolidine-2-carboxylic acid, MT2C) were synthesized and isolated by a literature protocol [12]. Figure 5.3 shows the results of a ^{1}H NMR experiment that assessed their formation and decomposition rates at pD 7 and pD 2 in D_2O at 25°C. It was found that, at pD 7, these compounds were stable over a period of two hours. The formation of these thioaminals from equimolar concentrations of cysteamine and α-oxoacid at pD 7 was also assessed. T2C forms rapidly, reaching equilibrium within 40 minutes. MT2C forms more slowly, presumably due to steric effects. The rates of formation of both these compounds were diminished at pD 2, presumably due to protonation of the amine nucleophile.

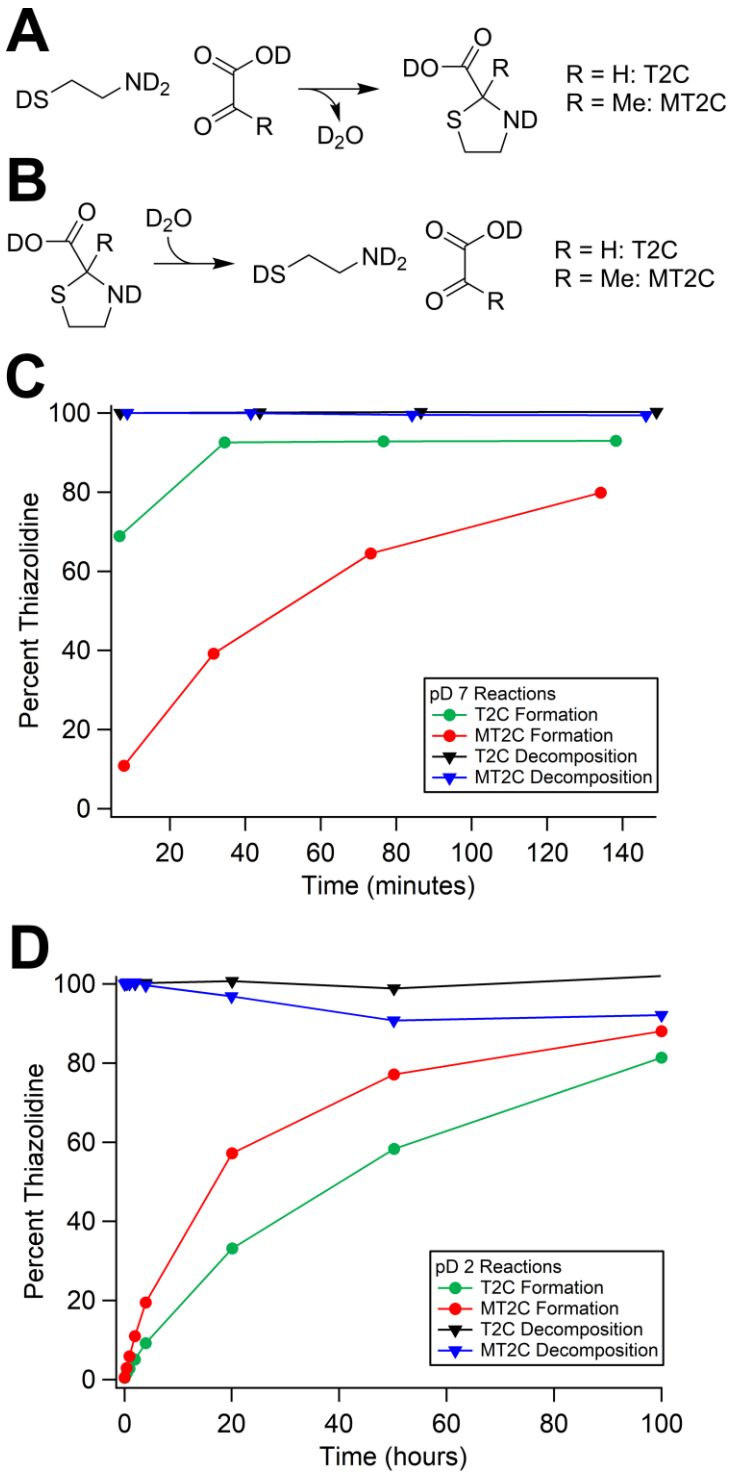


Figure 5.3. Formation and decomposition of thiazolidines (cyclic thioaminals) in D_2O. A. Formation reaction of thiazolidines. B. Decomposition reaction of thiazolidines. C. Extents of formation and decomposition reactions of thiazolidines in D_2O at pD 7 measured in percent molar composition. D. Extents of formation and decomposition reactions of thiazolidines in D_2O at pD 2 measured in percent molar composition.

Because a thioaminal backbone would require the formation of a linear, rather than a cyclic, thioaminal, an attempt was made to synthesize such a model compound. Employing the same method used for the synthesis of the cyclic thioaminal compounds, an attempt was made to synthesize and isolate a linear thioaminal compound from glyoxylic acid, ethanolamine, and β-mercaptoethanol. A white solid was isolated and was analyzed by ^{1}H NMR (Figure 5.4). Although the spectrum superficially resembled the intended product, spiking with ethanolamine did not produce any new peaks, suggesting that the isolated product was the ethanolammonium hemithioacetate carboxylate salt.

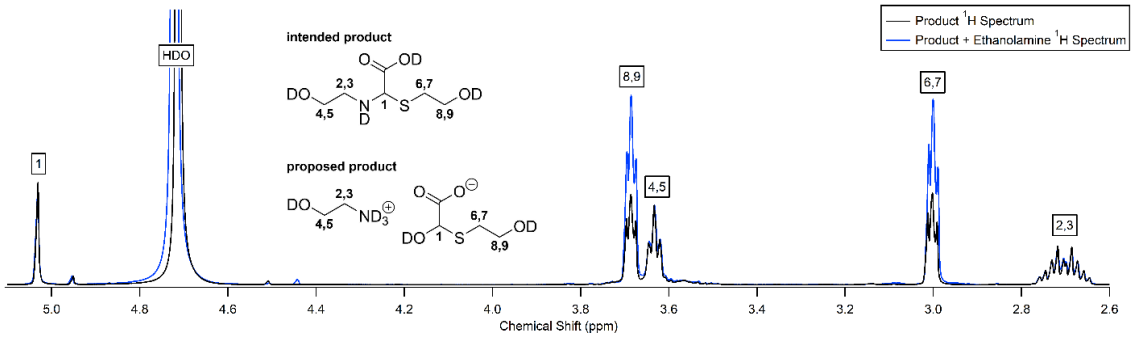


Figure 5.4. ^{1}H NMR analysis of the product obtained from the reaction of glyoxylic acid, ethanolamine, and β-mercaptoethanol. Although the spectrum is superficially consistent with the intended product, spiking the sample with ethanolamine reveals that the peaks assigned to the amino substituent of the product are actually free ethanolamine. Due to the presence of diastereotopic peaks (2,3) for the sulfide substituent, a hemithioacetal product is proposed.

Despite this negative result, the synthesis of a model proto-nucleic acid monomer featuring amino and mercepto moieties was investigated. It was hypothesized that, in an unassembled state, a monomer containing amino and mercapto moieties separated by 2 or 3 carbon atoms would react with glyoxylate to predominantly form a 5-membered or 6-membered (respectively) cyclic thioaminal, but that, once assembled, this cyclic form would equilibrate with a linear polymer with a 5-atom or 6-atom repeat unit (respectively).

Since any self-assembling nucleobase chosen would exhibit nucleophilic properties, an electrophile was required that would produce a nucleobase-functionalized amino thiol. Homocysteine thiolactone was chosen since nucleophilic attack of the thioester moiety would release a merapto group. Homocysteine thiolactone, the cyclic form of the amino acid homocysteine, is also conceivably prebiotic, as homocysteine can be synthesized from the addition of hydrogen sulfide to acrolein, followed by the typical Strecker synthesis [13]. Cyclization by thioesterification follows under acidic conditions.

It was found that 2,4,5,6-tetraaminopyrimidine (TetAP) successfully reacted with DL-*N*-acetylhomocysteine thiolactone in a solvent-free melt reaction. Without intermediate purification, the product of this reaction could be cyclized to the corresponding 8-substituted 2,6-diaminopurine by heated the crude reaction mixture in aqueous sodium hydroxide, which also simultaneously removed the *N*-acetyl group. This synthesis is depicted in Figure 5.5. The product, 1-(2,6-diaminopurin-8-yl)-homocysteamine (DAPHC), could be purified chromatographically.

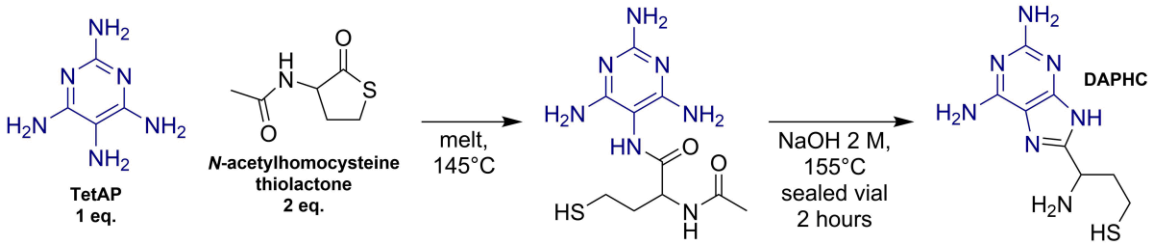


Figure 5.5. The synthesis of model proto-nucleic acid monomer DAPHC.

Unlike any self-assembling molecules previously synthesized, DAPHC features a 2,6-diaminopurine recognition unit. This unit has the same donor-acceptor-donor hydrogen-bonding pattern as TAP and melamine, and with a pK_a of 5, is expected to pair with either cyanuric acid or barbituric acid (refer to Figure 2.1). Hydrogelation of DAPHC

with cyanuric acid did not occur at any pH between 5 (the pK_a of diaminopurine) and 7 (the pK_a of cyanuric acid). However, it was reasoned that if the molecule could be endowed with a peripheral negative charge, hydrogelation would occur. When DAPHC was combined with a buffer containing cyanuric acid and either glyoxylic acid or pyruvic acid at pH 6, hydrogelation occurred, presumably due to formation of cyclic thioaminals with these α-oxoacids.

Mass spectrometry was performed on these hydrogels to test the original hypothesis that cyclic thioaminals would exist in some equilibrium with linear thioaminals. However, no oligomeric species could be detected (Figure 5.6). When the gels were analyzed by ^1H NMR, it appeared that DAPHC had only been converted to the cyclic thioaminal.

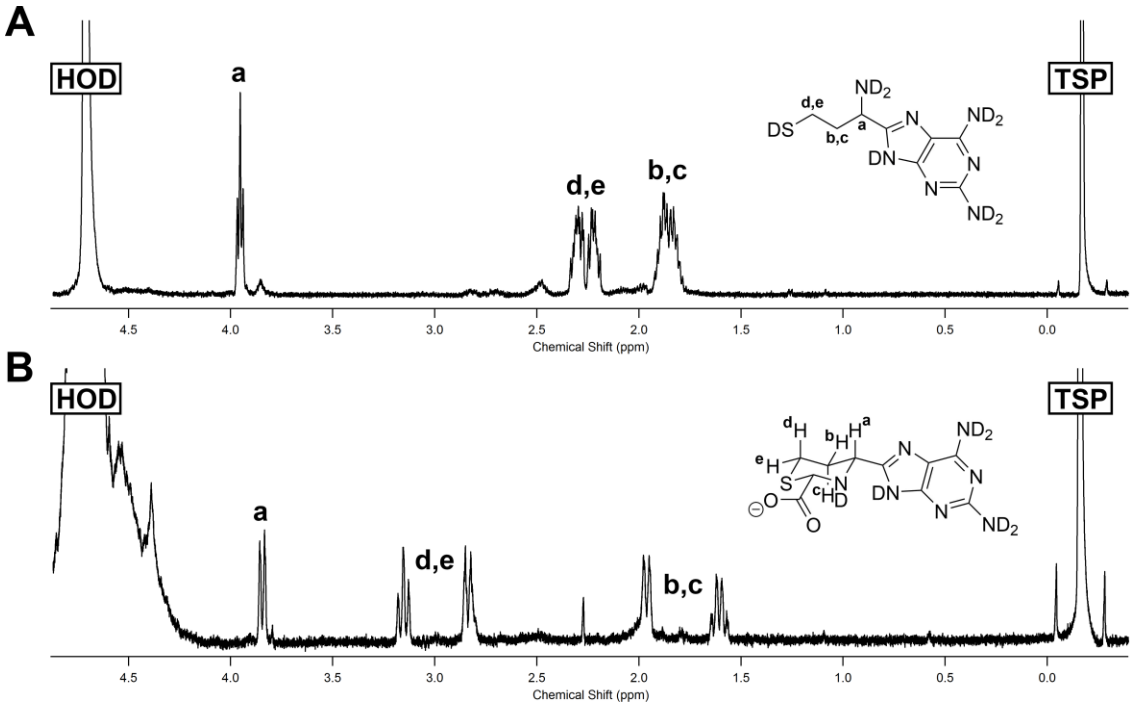


Figure 5.6. ^{1}H NMR Analysis of DAPHC in D_2O. A. ^{1}H spectrum of the parent compound DAPHC, with proton assignments. B. ^{1}H spectrum of the gel formed from DAPHC, glyoxylic acid, and cyanuric acid. Only a single species can be detected. As the cyclic thioaminal is expected to adopt a chair-like conformation, the presence of a single product may indicate that the diastereomer in which both substituents are arranged axially is produced exclusively. Note that, although a single enantiomer is depicted, the compound is present as a racemic mixture.

Although a thioaminal polymer could not be generated from this system, an important principle in polymer synthesis was solidified: in a polymer formed from condensation reactions that has a 5-atom or 6-atom backbone repeat length may also be prone to cyclization, and if the cyclic product is not thermodynamically destabilized by, for example, ring strain or electronic effects, then it will predominate over the linear polymer. In light of this result, new systems were investigated that would not suffer from the formation of stable cyclic monomers.

5.4 The Ester as the Chemical Linkage in the First Genetic Polymers

In the matrix of possible chemical linkages formed by prebiotic nucleophiles and electrophiles in Figure 1.5, the ester stands out as a functional group that can be easily formed in a plausibly prebiotic manner from dry-down reactions of alcohols and carboxylic acids under acidic conditions [14]. The acidic nature of the reaction medium is usually from the carboxylic acid itself, and not from an externally added acid. In prebiotic chemistry, the ester has primarily been investigated in the context of proto-polypeptide formation [15]. The ester has been proposed as a precursor to the amide linkage found in proteins, as the ribosome is also capable of forming esters [16]. Weber investigated the prebiotic formation of polyesters by the polymerization of glyceric acid, the α-hydroxy acid analogue of serine, under plausibly prebiotic conditions, and detected oligomer formation [17]. Hud and coworkers investigated the polymerization of malic acid, the α-hydroxy acid analogue of aspartic acid, and found that up to 52% of the total malic acid could be converted to trimers or higher [14].

As mentioned in § 4.5, malic acid, the hydroxy acid analogue of aspartic acid, is a particularly interesting example because it contains one nucleophilic site (a hydroxyl group) and two electrophilic sites (two carboxylic acids). Because of this, it has the potential to oligomerize by its nucleophilic site and one of its electrophilic sites, and to append nucleophilic nucleobases to itself at the remaining electrophilic site. The hydroxy acid analogue of glutamic acid (2-hydroxyglutaric acid) is expected to behave similarly. However, this compound, unlike malic acid, would cyclize to form a lactone (butyrolactone-4-carboxylic acid, BL4C).

Because 5-aminopyrimidines such as TetAP have the ability to react with esters, and because BL4C contains a lactone (i.e., a cyclic ester) and malic acid readily forms esters from dry-down reactions, a reaction that contains either malic and or BL4C and a 5-aminopyrimidine, such as TetAP or 5AU (in the appropriate ratio), might form oligoesters that are functionalized with a self-assembling noncanonical nucleobase. Although these oligomers would probably be irregular in structure, a successful reaction would demonstrate the viability of oligoesters in the context of proto-nucleic acid formation. Additionally, although this system may also produce cyclic hydroxy acid dimer products, these products, like the parent compound glycolide (1,4-dioxane-2,5-dione), are not a thermodynamic product; rather, they tend to act as substrates for ring-opening polymerization.

Reactions were conducted in which aqueous solutions of either L-malic acid or (*S*)-BL4C and either TetAP or 5AU, of variable molar ratios, were cycled by drying at 100°C and rehydrated after 24 hours. After three cycles, the reactions were analyzed by mass spectrometry with electrospray ionization (ESI-MS), a soft ionization technique that typically shows molecular weights plus or minus a proton (in positive or negative ion mode, respectively). Some results are shown in Figure 5.7 for reactions performed with a ratio of nucleobase-to-monomer of 1:4 after three wet-dry cycles at 100°C.

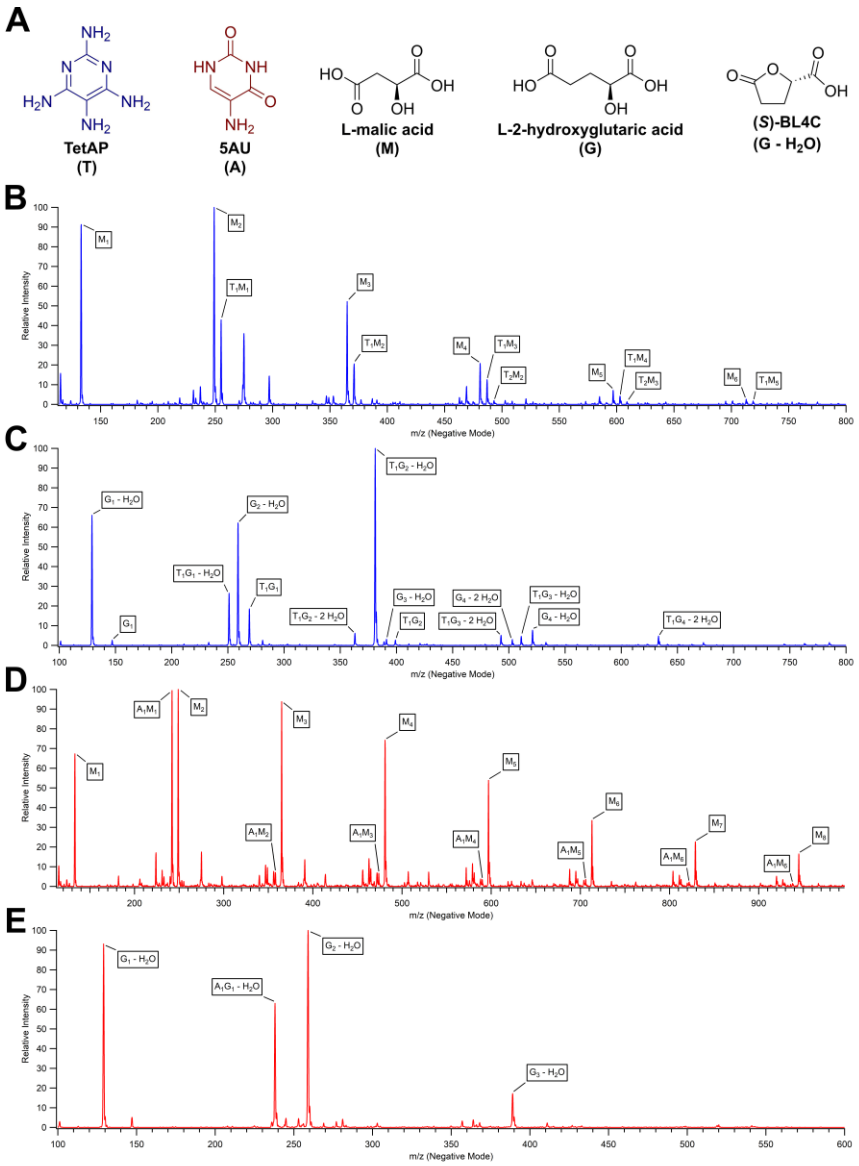


Figure 5.7. Annotated ESI-MS spectra of the oligomerization reactions of either TetAP or 5AU with either L-malic acid or (*S*)-BL4C (the lactone form of L-2-hydroxyglutaric acid) in a 1:4 molar ratio after three 24-hour wet-dry cycles at 100°C. The annotation is shown in the format X_aY_b, where X = T (TetAP), A (5AU) and Y = M (malic acid), G (2-hydroxyglutaric acid). The mass of the species is given by $m/z = am_X + bm_Y - (a + b - 1)m_{H_2O}$. When water losses are shown explicitly, these correspond to further losses after those already accounted for by the given formula. A. Structures of the compounds used in oligomerization reactions with one-letter abbreviations. B. Mass spectrum of the oligomerization of malic acid with TetAP. C. Mass spectrum of the oligomerization of BL4C with TetAP. D. Mass spectrum of the oligomerization of malic acid with 5AU. E. Mass spectrum of the oligomerization of BL4C with 5AU.

Malic acid oligomerizes more efficiently than BL4C regardless of the 5-aminopyrimidine added. This is perhaps unsurprising as malic acid features a nucleophilic hydroxyl group, whereas BL4C must be hydrolyzed to reveal a nucleophilic hydroxyl group. TetAP is more efficiently incorporated into the oligomers than 5AU regardless of the monomer added. Again, this is unsurprising as TetAP is more electron-rich than 5AU, making it more nucleophilic.

The success of this system motivated further study into nucleobase-functionalized oligoesters. A synthesis of a cyanuric acid-functionalized hydroxy acid was attempted based on the known S_N2-type reactivity of cyanuric acid in other conventional synthetic organic reactions [18]. This synthesis involved the alkylation of cyanuric acid with α-hydroxy-γ-butyrolactone (HBL) in DMF with a base catalyst (Figure 5.8). The reaction was successful in producing a 2-hydroxy-4-cyanuryl-butanoic acid (a cyanuric acid-functionalized hydroxy acid, Cy^{HA}).

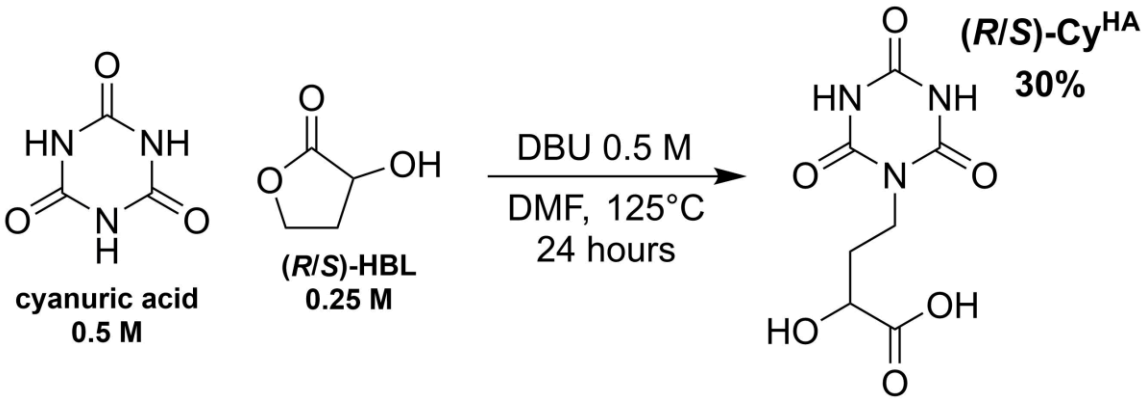


Figure 5.8. The synthesis of racemic Cy^{HA}, a cyanuric acid-functionalized hydroxy acid.

The co-oligomerization of this compound with racemic malic acid was attempted. When an aqueous solution of Cy^{HA} (free acid) and malic acid in a 1:1 ratio (pH ~2) was

cycled by drying at 85°C for 24 hours and rehydrating. The reaction was then analyzed by ESI-MS. After only two cycles, a significant yield of oligomeric products containing both Cy^{HA} and malic acid was found (Figure 5.9).

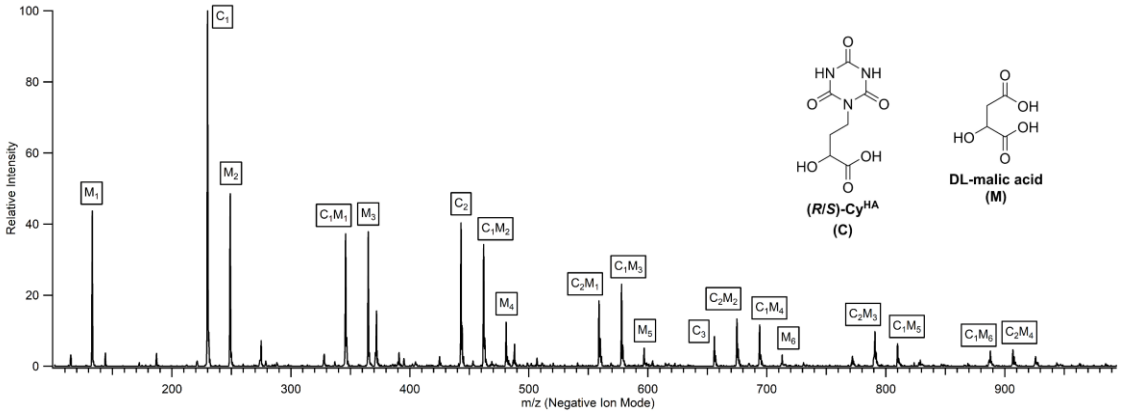


Figure 5.9. Annotated ESI-MS spectrum of the oligomerization reaction of Cy^{HA} with malic acid in a 1:1 molar ratio after two 24-hour wet-dry cycles at 85°C. The annotation is shown in the format C_aM_b, with the mass of the species is given by $m/z = am_C + bm_M - (a + b - 1)m_{H_2O}$.

Because of the success of this oligomerization reaction, the properties of Cy^{HA} were studied in greater detail. The pK_a value of the cyanuric acid moiety of Cy^{HA} was determined by a ^1H NMR titration experiment and found to be, like cyanuric acid, close to neutrality at 7.42. The self-assembly properties of this compound were tested cursorily with melamine, TAP, and TetAP, by dissolving both components in water at 50 mM at room temperature and at a given pH, which was adjusted with NaOH or HCl. It was found that Cy^{HA} formed a stable hydrogel with TetAP at pH 7. With melamine a stable hydrogel could not be formed. Instead, either at pH 6 or pH 8, a white precipitate formed. With TAP, Cy^{HA} could formed white precipitates at pH 7 and 7.5, but a stable hydrogel could be formed at pH 8.

Previous studies on TAP-cyanuric acid supramolecular assemblies in water found that, below a certain critical concentration, no self-assembly occurs. Above this critical concentration, called the minimum assembly concentration (MAC), rapid supramolecular polymerization occurs to form stacked-hexad fibers [18]. When analyzed by ^{1}H NMR, the molecules present in the supramolecular assembly tumble so slowly that their signals broaden to baseline. The result of this is that only the molecules that remain free in solution can be detected, and their measured concentration is equal to the MAC. By preparing a sample of Cy^{HA} and TAP at pH 8 (with a sodium ethylphosphonate buffer) at some concentration above the MAC and incrementally diluting the sample, a plot can be formed of NMR-measured concentration versus actual concentration that reveals the MAC value as a plateau. This experiment was performed and it was found that Cy^{HA} and TAP, at pH 8, self-assemble with an MAC of 30 mM (Figure 5.10).

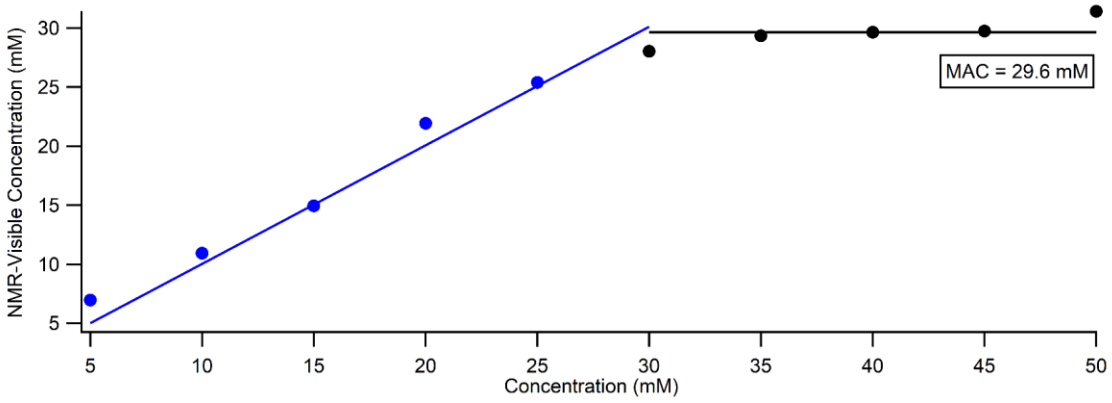


Figure 5.10. Minimum assembly concentration (MAC) determination between Cy^{HA} and TAP at pD 8 in D_2O. Upon supramolecular assembly, the components become undetectable by NMR, and only the amount that remains unassembled is measurable. This measured concentration of approximately 30 mM corresponds to the MAC.

5.5 Conclusions

The thioaminal system studied here was ultimately found to be unsuitable for prebiotic polymer formation. Cyclic thioaminals with either 5 or 6 ring members were found to form readily, but were not found to equilibrate with linear thioaminals. Thermodynamically, this may be rationalized by supposing that the formation of cyclic and linear thioaminals is approximately equally enthalpically favorable, but that the formation of a cyclic thioaminal is entropically favored over linear thioaminal formation. For a compound such as cysteamine that contains amino and mercapto moieties, after initial imine formation, intramolecular attack by a thiol(ate) moiety is extremely fast and concentration-independent. However, the formation of linear thioaminals would clearly have a concentration dependence in both the amine nucleophile and the thiol nucleophile.

In order to circumvent this problem of cyclization competing with polymerization, a system must be chosen in which the cyclic form, which, for most cases, will be entropically favored over a linear form, is enthalpically disfavored. Enthalpic destabilization of a cyclic form can come from steric and electronic factors. For example, thioaminal polymers can be formed when the monomeric constituents are incapable of forming stable 5-membered or 6-membered rings [19].

In the case of esterification, oligomerzation of α-hydroxy acids has the potential to form cyclic 1,4-dioxane-2,5-dione dimers (i.e., derivatives of glycolide). Glycolide and its derivatives are often used to form polyesters [20]. The successful polymerization of glycolide is due to ring strain [21], as the sp^2 hybridization of four of its endocyclic atoms prevents a chair-like conformation. Instead, it forms a strained boat-like conformation [22].

This instability of the 1,4-dioxane-2,5-dione (i.e., cyclic hydroxy acid dimers) is necessary for the prebiotic viability of hydroxy acid oligomers.

Although oligoesters that are appended with self-assembling noncanonical nucleobases can be formed, the systems studied here have several limitations. First, the formation of an informational system requires many recognition units so that large amounts of information can be stored and replicated. In the systems studied here, oligomers that contained more than two recognition units were not detected. Second, complementary self-assembly is necessary for template-directed replication of proto-nucleic acids. The oligomers formed in these systems were probably of irregular geometry, which has the potential to complicate self-assembly. Third, the ester is liable to hydrolysis under certain conditions, and it is not clear whether it is stable enough to support an informational system. As will be explained in Chapter 6, small modifications to the oligoester systems studied here may ameliorate these problems and enhance the prebiotic viability of a proto-nucleic acid formed by oligoesterification.

5.6 References

[1] Shapiro, R., Prebiotic ribose synthesis: a critical analysis. Orig. Life Evol. Biosph., 1988. 18: p. 71-85.

[2] Decker, P., P. Schweer, and R. Pohlmann, Identification of formose sugars, presumable prebiotic metabolites, using capillary gas chromatography/gas chromatography-mass spectroscopy of n-butoxime trifluoroacetates on OV-225. J. Chromatogr., 1982. 244: p. 281-291.

[3] Schwartz, A.W., Phosphorus in prebiotic chemistry. Philosophical transactions of the Royal Society of London, Series B: Biological Sciences, 2006: p. 1743-1749.

[4] Kim, H.J., et al., Synthesis of carbohydrates in mineral-guided prebiotic cycles. Journal of the American Chemical Society, 2011. 133(24): p. 9457-9468.

[5] Ricardo, A., et al., Borate minerals stabilize ribose. Science, 2004. 303(5655): p. 196-196.

[6] Gibard, C., et al., Phosphorylation, oligomerization and self-assembly in water under potential prebiotic conditions. Nat. Chem., 2018. 10(2): p. 212-217.

[7] Burcar, B., et al., Darwin's Warm Little Pond: A one-pot reaction for prebiotic phosphorylation and the mobilization of phosphate from minerals in a urea-based solvent. Angew. Chem. Int. Ed. Engl., 2016. 55: p. 13249-13253.

[8] Weber, A.L., The sugar model: Catalysis by amines and amino acid products. Orig. Life Evol. B., 2001. 31: p. 71-86.

[9] Hud, N.V., et al., The origin of RNA and 'My Grandfather's Axe'. Chem. Biol., 2013. 20: p. 466-474.

[10] Hud, N.V. and F.A.L. Anet, Intercalation-mediated synthesis and replication: a new approach to the origin of life. Journal of Theoretical Biology, 2000. 205(4): p. 543-562.

[11] Bean, H.D., et al., Glyoxylate as a backbone linkage for a prebiotic ancestor of RNA. Origins of Life and Evolution of the Biosphere, 2006. 36(1): p. 39-63.

[12] Lalezari, I. and E.L. Schwartz, Synthesis and antineoplastic activity of 5-aryl-2,3-dihydropyrrolo[2,1-b]thiazole-6,7-dimethanol 6,7-bis(isopropylcarbamates). J. Med. Chem., 1988. 31(7): p. 1427-9.

[13] Van Trump, J.E. and S.L. Miller, Prebiotic synthesis of methionine. Science, 1972. 178(4063): p. 859-60.

[14] Mamajanov, I., et al., Ester formation and hydrolysis during wet-dry cycles: generation of far-from-equilibrium polymers in a model prebiotic reaction. Macromolecules, 2014. 47: p. 1334-1343.

[15] Forsythe, J.G., et al., Ester-Mediated Amide Bond Formation Driven by Wet-Dry Cycles: A Possible Path to Polypeptides on the Prebiotic Earth. Angewandte Chemie-International Edition, 2015. 54(34): p. 9871-9875.

[16] Fahnestock, S. and A. Rich, Ribosome-catalyzed polyester formation. Science, 1971. 173(3994): p. 340-343.

[17] Weber, A.L., Thermal synthesis and hydrolysis of polyglyceric acid. Origins of Life and Evolution of the Biosphere, 1989. 19(1): p. 7-19.

[18] Cafferty, B.J., et al., Ultra-sensitive pH control of supramolecular polymers and hydrogels: pK(a) matching of biomimetic monomers. Chemical Science, 2014. 5(12): p. 4681-4686.

[19] Wojtecki, R.J., et al., Developments in Dynamic Covalent Chemistries from the Reaction of Thiols with Hexahydrotriazines. J. Am. Chem. Soc., 2015. 137(45): p. 14248-51.

[20] Chujo, K., et al., Ring-opening polymerization of glycolide. Die Makromolekulare Chemie, 1967. 100(1): p. 262-266.

[21] Saiyasombat, W., et al., Ring strain and polymerizability of cyclic esters. Polymer, 1998. 39(23): p. 5581-5585.

[22] Belen'kaya, B.G., et al., Crystal and molecular structures of glycolide and lactide: Association through CH···O hydrogen bonds. Crystallography Reports, 1997. 42(3): p. 403-406.

CHAPTER 6. DEPSIPEPTIDE NUCLEIC ACIDS

6.1 Introduction

The prebiotic emergence of the first genetic polymers is a poorly understood phenomenon [1]. The influential RNA world hypothesis suggests that pre-formed nucleosides or nucleotides, through some hitherto unclear mechanism, spontaneously polymerized to form RNA [2]. However, the formation of intact oligomeric RNA from mononucleotides in a plausibly prebiotic manner is challenging. This, and other issues associated with the initial formation of the nucleosides (see § 1.2.1), suggest that RNA is a product of evolution, and was preceded by a proto-RNA polymer that was similar in function, but more easily accessed in a prebiotic manner [3, 4].

In the study of proto-RNA, a number of alternative nucleobases have been suggested that are more reactive with ribose [5-10] and other sugars [11, 12] than the canonical nucleobases, and have greater propensities for self-assembly [13]. Similarly, it may be that the features of the chemical backbone of proto-RNA differed significantly from RNA. The chemical backbone of RNA consists of two principle components: the ribose sugar and the phosphate linker. A number of studies have been performed on the formation and base-pairing properties of nucleic acids with alternative sugars [14, 15]. Studies have also been conducted on the formation of the phosphodiester linkage [16-18]. Despite these proposed prebiotic syntheses, the prebiotic feasibility of the phosphodiester has been disputed [19]. A number of prebiotic alternatives have been suggested, such as acetals [20], amides [21, 22], phosphite esters [23], or thiophosphate esters [24]. The ester linkage has been investigated for nucleic acid analogues, but has only been formed under

prebiotically unrealistic conditions with a carbodiimide condensing agent [25]. In the context of proto-polypeptide formation, it has been found that the ester is formed readily under relatively mild drying conditions [26]. It may be that proto-nucleic acids were also formed from ester linkages, which were, through a process of chemical and/or biological evolution, eventually replaced by the more optimal, but prebiotically more difficult to access, phosphodiester.

In this study, a new class of plausibly prebiotic genetic polymer is described that is based on nucleobase-functionalized α-hydroxy acid-α-amino acid heterodimers as the candidate proto-nucleic acid monomers. These model proto-nucleic acid monomers oligomerize by esterification to form depsipeptide nucleic acids. The propensities for self-assembly and resistance to hydrolysis of these oligomers were determined to evaluate whether or not this new class polymer is viable as an early informational system.

6.2 Experimental Methods

6.2.1 Materials

Cyanuric acid and melamine were purchased from Acros Organics and used as received. 2-Mercaptopyridine was purchased from Millipore Sigma and recrystallized from benzene prior to use. All other compounds were purchased from Millipore Sigma and used as received.

6.2.2 Analytical Methods

NMR spectra were collected on a Bruker DRX 700 MHz NMR spectrometer or a Bruker DRX 800 MHz NMR spectrometer and were the sum of 64 transients (^{1}H) or 1024 transients (^{13}C).

AFM images were obtained with a Nanoscope IIIa (Digital Instruments) in tapping mode using silicon tips (Vistaprobes, 48 N/m) on freshly cleaved mica. Samples were deposited onto mica in 3 μL aliquots, incubated in a humid atmosphere for 10 minutes, and then blown dry with a stream of nitrogen.

UV-LC/MS analysis was performed with a method with the following specifications:

Column: Waters XBridge Amide, 3.5 μm, 2.1 x 150 mm
Column Temperature: 25°C
Flow Rate: 0.5 mL/min
Injection Volume: 1 μL
Mobile Phase A: Aqueous NH_4HCO_2 buffer, 10 mM, pH 9
Mobile Phase B: MeCN
Gradient:

Time (min)	%A	%B
0	10	90
4	25	75
11	35	65
11.2	40	60
12.8	40	60
13	10	90
18	10	90

6.2.3 Synthesis

(*R*/*S*)-Cy^{HA}: To 100 mL of DMF, 6.45 g of cyanuric acid (50 mmol) and 7.48 mL of DBU (50 mmol) were added. The mixture was heated to 125°C with rapid stirring. Once the cyanuric acid had fully dissolved, 1.95 mL of DL-α-hydroxy-γ-butyrolactone (25 mmol) were added dropwise. The reaction was stirred at 125°C for 24 hours. The solution was then poured onto a watch glass and allowed to evaporate over several days, eventually producing a thick, brown oil above crystals of unreacted cyanuric acid. This material was suspended in 50 mL of water, and the suspension was centrifuged. The pellet was discarded and the supernatant was separated, in 5 mL portions, using a Teledyne Isco CombiFlash Rf+ purification system with a RediSep C18Aq 150g Gold column with a 100% water eluent. The fractions containing product, as detected by LC/MS, were pooled and reduced to about 10 mL with a rotary evaporator. To remove residual DBU, this solution was eluted over 25 g of Dowex 50WX8 cation exchange resin, hydrogen form, with 0.1 % aqueous formic acid solution. 100 mL were collected and the solvent was removed with a rotary evaporator to give white flakes of racemic compound Cy^{HA} in its free acid form (1.73 g, 30%). ^{1}H NMR (700 MHz, D_2O): δ 4.33 (dd, 8.1 Hz, 4.1 Hz, 1H, H^2), 3.96 (m, 2H, H^4), 2.15 (m, 1H, H^3), 2.01 (m, 1H, $H^{3'}$). ^{13}C NMR (176 MHz, D_2O): δ 177.2 (C1), 150.9 (C5), 150.0 (C6), 68.1 (C2), 37.9 (C4), 31.0 (C3). HRMS (m/z): $[M+H]^+$ $C_7H_{10}O_6N_3$, theoretical mass: 232.0564, actual mass: 232.0568.

(*R*/*S*)-Cy^{HA} Isopropyl Ester: In a glass vial equipped with a stir bar, Cy^{HA} (116 mg, 0.5 mmol) was suspended in 1 mL isopropanol, and TMSCl (12.7 µL, 0.1 mmol) was added. The vial was sealed and the mixture was stirred rapidly at 100°C for six hours, producing a solution. The solvent was then evaporated first by blowing a stream of nitrogen onto the

solution, and then by removing any residual solvent under vacuum, giving racemic Cy^{HA} isopropyl ester as a white solid (137 mg, quant.). The solid was left in the glass vial and taken to the next synthetic step without any further purification. ^{1}H NMR (700 MHz, d6-DMSO): δ 11.37 (s, 2H, H^{b}), 5.43 (d, 5.5 Hz, 1H, H^{a}), 4.91 (m, 6.2 Hz, 1H, H^{7}), 4.04 (m, 1H, H^{2}), 3.74 (m, 2H, H^{4}), 1.89 (m, 1H, H^{3}), 1.76 (m, 1H, $H^{3'}$). ^{13}C NMR (176 MHz, d6-DMSO): δ 173.5 (C1), 150.7 (C5), 149.1 (C6), 68.7 (C2 or C7), 68.1 (C2 or C7), 38.0 (C4), 32.1 (C3), 22.0 (C8 or C9), 21.9 (C8 or C9). HRMS (m/z): $[M+H]^{+}$ $C_{10}H_{16}O_{6}N_{3}$, theoretical mass: 274.1034, actual mass: 274.1039.

(*S*,*S*)-Cy^{HA}-D: To a glass vial containing a recently prepared sample of racemic Cy^{HA} isopropyl ester (137 mg, 0.5 mmol), vacuum-dried L-aspartic acid (i.e., (*S*)-D, 83.2 mg, 625 µmol) and dry DBU (0.3 mL) were added. The glass vial, equipped with a stir bar, was sealed and stirred at 100°C for 48 hours, forming a thick, brown solution. Upon returning to room temperature, the solution formed a thick gel. This gel was dissolved in 450 µL water with 50 µL formic acid, and the pH of the resulting solution was adjusted to 5 by the further addition of formic acid. The solution was purified in three chromatographic stages. In the first stage, the solution was loaded onto a column containing 75 mL of QAE Sephadex A-25 resin equilibrated in ammonium formate buffer, 50 mM, pH 5, and separated using a Teledyne Isco CombiFlash Rf+ purification system with a gradient of ammonium formate pH 5, 50 mM, to ammonium formate, pH 5, 500 mM. The fractions containing product were pooled and reduced to about 5 mL with a rotary evaporator. In the second stage of purification, this solution was eluted over 10 g of Dowex 50WX8 cation exchange resin, hydrogen form, with 0.1 % aqueous formic acid solution. 100 mL were collected and the solvent was reduced to about 5 mL with a rotary evaporator. In the third

stage of purification, this solution was separated using a Teledyne Isco CombiFlash Rf+ purification system with a RediSep C18Aq 150g Gold column with a 100% water eluent. This method of purification gave a slight degree of separation between the two diastereomers of Cy^{HA}-D. The fractions were then separated into three groups: the first containing pure diastereomer 1 (i.e., the first diastereomer to elute), the second containing a mixture of the two diastereomers, and the third containing pure diastereomer 2 (i.e., the second diastereomer to elute). The solvent was removed from each using a rotary evaporator, and the second group (containing a mixture of the two diastereomers) was again separated using a Teledyne Isco CombiFlash Rf+ purification system with a RediSep C18Aq 150g Gold column with a 100% water eluent. This process was continued until the diastereomers were fully separated. 25.6 mg of the (*S*,*S*)-diastereomer (14.8%) and 29.2 mg of the (*R*,*S*)-diastereomer (16.9%) were obtained.

(*S*,*S*)-Cy^{HA}-D: ^{1}H NMR (800 MHz, D_2O, 5°C): δ 4.58 (dd, 6.7 Hz, 5.1 Hz, 1H, H^8), 4.05 (dd, 8.4 Hz, 3.6 Hz, 1H, H^2), 3.75 (m, 1H, H^4), 3.68 (m, 1H, $H^{4'}$), 2.77 (m, 2H, H^9), 1.89 (m, 1H, H^3), 1.69 (m, 1H, $H^{3'}$). ^{13}C NMR (201 MHz, D_2O, 5°C): 175.6 (C1, C7, or C10), 174.3 (C1, C7, or C10), 173.8 (C1, C7, or C10), 150.7 (C5), 149.9 (C6), 68.9 (C2), 48.5 (C8), 37.5 (C4), 35.2 (C9), 31.0 (C3). HRMS (m/z): $[M+H]^+$ $C_{11}H_{15}O_9N_4$, theoretical mass: 347.0834, actual mass: 347.0840.

(*R*,*S*)-Cy^{HA}-D: ^{1}H NMR (800 MHz, D_2O, 5°C): δ δ 4.57 (dd, 6.8 Hz, 5.1 Hz, 1H, H^8), 4.03 (dd, 8.4 Hz, 3.6 Hz, 1H, H^2), 3.75 (m, 1H, H^4), 3.70 (m, 1H, $H^{4'}$), 2.77 (m, 2H, H^9), 1.88 (m, 1H, H^3), 1.72 (m, 1H, $H^{3'}$). ^{13}C NMR (201 MHz, D_2O, 5°C): 175.6 (C1, C7, or C10), 174.4 (C1, C7, or C10), 173.7 (C1, C7, or C10), 150.7 (C5), 149.9 (C6), 68.9 (C2), 48.6 (C8), 37.5 (C4), 35.2 (C9), 31.1 (C3).

(Cy^{HA}-D)$_2$: 51.9 mg (*S*,*S*)-Cy^{HA}-D (150 µmol) were dissolved in 1.2 mL water. 47 µL of pyrimidine (48 mg, 600 µmol) were added to the solution. To six Fisherfinest Premium Superslip cover glasses (each 2.4 cm x 4 cm) that were previously desilanized by heating in a bath of 5 M NaOH at approximately 90°C for two hours and washed with deionized water, the solution of ψD and pyrimidine was applied evenly (200 µL per cover glass) across the full surface area. These cover glasses were placed uncovered in an oven at 85°C for two weeks. After this period, the material deposited on the cover glasses was recovered by redissolving the material on each slide with 200 µL water, followed by two further washings with 100 µL water. The material from all six slides was pooled. The components of the solution were then separated by preparative HPLC using an Agilent Prep-C18 column 50 mm x 21.2 mm. The solution was injected onto the column in 300 µL portions and eluted with a gradient of 98% aqueous/2% acetonitrile for 5 minutes, followed by a ramp to 85% aqueous/15% acetonitrile over 12 minutes. The aqueous phase used was 0.1% TFA. Four closely eluting peaks with the mass of (Cy^{HA}-D)$_2$ (two major peaks and two minor peaks) were collected. The minor peaks were assumed to be epimerization products and were not used in this study. The two major peaks were lyophilized to yield 6.9 mg (13.6%) of one (Cy^{HA}-D)$_2$ isomer and 3.5 mg (6.9%) of another (Cy^{HA}-D)$_2$ isomer. Based on the chemical shift of the *β*-protons of the first aspartic acid residue of each dimer, it was found that the primary dimer (i.e., produced in higher yield) was *α*-(Cy^{HA}-D)$_2$ and the secondary dimer (i.e., produced in lower yield) was *β*-(Cy^{HA}-D)$_2$.

α-(Cy^{HA}-D)$_2$: ^{1}H NMR (800 MHz, D_2O): δ 5.16 (dd, 7.0 Hz, 5.2 Hz, 1H, H^{12}), 4.99 (t, 6.1 Hz, 1H, H^8), 4.77 (dd, 7.0 Hz, 5.2 Hz, 1H, H^{18}), 4.30 (dd, 8.6 Hz, 3.5 Hz, 1H, H^2), 4.01-3.87 (m, 4H, H^4, $H^{4'}$, H^{14}, $H^{14'}$), 3.09 (dd, 17.2 Hz, 5.6 Hz, 1H, H^9), 3.04 (dd, 17.2 Hz, 6.8

Hz, 1H, $H^{9'}$), 3.00 (dd, 17.1 Hz, 5.2 Hz, 1H, H^{19}), 2.96 (dd, 17.1 Hz, 7.0 Hz, 1H, $H^{19'}$), 2.21 (m, 2H, H^{13}, $H^{13'}$), 2.13 (m, 1H, H^{3}), 1.94 (m, 1H, $H^{3'}$). ^{13}C NMR (201 MHz, D_2O): 175.8 (C1), 174.2 (C20), 174.0 (C10), 173.5 (C17), 170.8 (C11), 170.7 (C7), 150.9 (C5), 150.8 (C15), 150.0 (C6 or C16), 149.9 (C6 or C16), 72.9 (C12), 69.3 (C2), 49.1 (C18), 48.7 (C8), 37.9 (C4), 37.3 (C14), 35.3 (C19), 35.1 (C9), 31.5 (C3), 28.9 (C13). HRMS (m/z): $[M+Na]^+$ $C_{22}H_{26}O_{17}N_8Na$, theoretical mass: 697.1308, actual mass: 697.1319.

β-(Cy^{HA}-D)$_2$: ^{1}H NMR (800 MHz, D_2O): δ 5.09 (dd, 8.4 Hz, 3.9 Hz, 1H, H^{12}), 4.87 (t, 6.0 Hz, 1H, H^{8}), 4.76 (dd, 6.9 Hz, 5.4 Hz, 1H, H^{18}), 4.29 (dd, 8.4 Hz, 3.6 Hz, 1H, H^{2}), 4.03-3.85 (m, 4H, H^{4}, $H^{4'}$, H^{14}, $H^{14'}$), 3.14 (m, 2H, H^{9}), 3.00 (dd, 17.2 Hz, 5.4 Hz, 2H, H^{19}), 2.96 (dd, 17.2 Hz, 6.9 Hz, $H^{19'}$), 2.18 (m, 2H, H^{13}, $H^{13'}$), 2.11 (m, 1H, H^{3}), 1.94 (m, 1H, $H^{3'}$). ^{13}C NMR (201 MHz, D_2O): δ 175.7 (C1), 174.2 (C20), 173.6 (C7), 171.3 (C17), 171.2 (C11), 150.9 (C5 or C15), 150.8 (C5 or C5), 150.0 (C6 or C16), 149.9 (C6 or C16), 72.4 (C12), 69.4 (C2), 49.0 (C18), 48.6 (C8), 37.9 (C4), 37.4 (C14), 35.4 (C9), 35.3 (C19), 31.3 (C3), 29.0 (C13). HRMS (m/z): $[M+Na]^+$ $C_{22}H_{26}O_{17}N_8Na$, theoretical mass: 697.1308, actual mass: 697.1306.

(*S*)-Mel^{HA}: To 225 mL of 1:1 EtOH:H_2O, 2-chloro-4,6-diamino-1,3,5-triazine (1.45 g, 10 mmol), (*S*)-4-amino-2-hydroxybutyric acid (1.79 g, 15 mmol), and sodium bicarbonate (1.68 g, 20 mmol) were added. The suspension was stirred at 80°C for two days, after which time it appeared as a clear solution. The solution was then transferred to an ice bath. Once the internal temperature of the solution had reached 0°C, concentrated HCl was added until the pH reacted 3.9. This produced a fine, white precipitate. The precipitate was filtered and washed with cold water and ethanol. The solid material was collected and dried under vacuum, giving compound (*S*)-Mel^{HA} (1.8 g, 79%). ^{1}H NMR (700 MHz, *d6*-DMSO): δ

6.48 (t, 5.6 Hz, 1H, H^c), 6.18 (s [br], 2H, H^d or H^e), 6.0 (s [br], 2H, H^d or H^e), 3.96 (dd, 8.8 Hz, 3.9 Hz, 1H, H^2), 3.28 (m, 2H, H^4), 1.87 (m, 1H, H^3), 1.63 (m, 1H, $H^{3'}$). ^{13}C NMR (176 MHz, d6-DMSO): δ 176.3 (C1), 167.3 (br, C6 or C7), 167.0 (br, C6 or C7), 166.6 (C5), 68.4 (C2), 37.1 (C4), 34.6 (C3). HRMS (m/z): $[M+H]^+$ $C_7H_{13}O_3N_6$, theoretical mass: 229.1044, actual mass: 229.1045.

(*S*)-Mel^{HA} Isopropyl Ester: In a glass vial equipped with a stir bar, compound (*S*)-Mel^{HA} (114 mg, 0.5 mmol) was suspended in 10 mL isopropanol, and TMSCl (635 µL, 5 mmol) was added. The vial was sealed and the mixture was stirred rapidly at 85°C overnight, producing a solution. The solvent was then evaporated first by blowing a stream of nitrogen onto the solution, and then by removing any residual solvent under vacuum, giving (*S*)-Mel^{HA} isopropyl ester hydrochloride salt as a white solid. The yield, as determined by ^{1}H NMR, was 84%. The solid was left in the glass vial and taken to the next synthetic step without any further purification. ^{1}H NMR (800 MHz, d6-DMSO): δ 8.14 (s [br], 1H, H^b), 8.04 (s [br], 2H, H^c or H^d), 7.83 (s [br], 2H, H^c or H^d), 4.91 (m, 6.3 Hz, 1H, H^8), 4.05 (dd, 8.9 Hz, 3.9 Hz, 1H, H^2), 3.38 (m, 2H, H^4), 1.89 (m, 1H, H^3), 1.72 (m, 1H, $H^{3'}$), 1.20 (d, 6.3 Hz, 3H, H^9 or H^{10}), 1.19 (d, 6.3 Hz, 3H, H^9 or H^{10}). ^{13}C NMR (201 MHz, d6-DMSO): δ 173.8 (C1), 158.8 (C6 or C7), 158.6 (C6 or C7), 157.0 (C5), 68.3 (C2 or C8), 68.1 (C2 or C8), 37.5 (C4), 33.6 (C3), 22.1 (C9 or C10), 22.0 (C9 or C10). HRMS (m/z): $[M+H]^+$ $C_{10}H_{19}O_3N_6$, theoretical mass: 271.1513, actual mass: 271.1517.

(*S,S*)-Mel^{HA}-D: To a glass vial containing a recently prepared sample of (*S*)-Mel^{HA} isopropyl ester hydrochloride (0.5 mmol), vacuum-dried L-aspartic acid (i.e., (*S*)-D, 83.2 mg, 625 µmol) and dry DBU (0.3 mL) were added. The glass vial, equipped with a stir bar, was sealed and stirred at 100°C for 48 hours, forming a brown precipitate. This mixture

was dissolved in 450 µL water with 50 µL formic acid, and the pH of the resulting solution was adjusted to 5 by the further addition of formic acid. The solution was loaded onto a column containing 75 mL of QAE Sephadex A-25 resin equilibrated in ammonium formate buffer, 50 mM, pH 5, and separated using a Teledyne Isco CombiFlash Rf+ purification system with a gradient of ammonium formate pH 5, 50 mM, to ammonium formate, pH 5, 500 mM. The fractions containing product were pooled and reduced to about 5 mL with a rotary evaporator. This solution was transferred to a conical tube and brought to a total volume of 12 mL by the addition of water. 12 mL of ethyl acetate were then added, and the solution was shaken vigorously for several minutes. After standing for about an hour, a bed of acetamide formed at the bottom of the aqueous layer. The organic phase was removed and the aqueous phase was heated to bring all of the acetamide into solution. This solution was separated using a Teledyne Isco CombiFlash Rf+ purification system with a RediSep C18Aq 150g Gold column with a 100% water eluent. The solution was injected onto the column in 4 mL portions. A small amount of another diastereomer (produced by epimerization during the synthesis) could be separated from the main product using this chromatographic method. Fractions containing the main product were pooled and reduced to about 5 mL with a rotary evaporator. This solution was then lyophilized to give the product, free of any counter-ions or buffer salts. ^{1}H NMR (800 MHz, D_2O): δ 4.51 (dd, 6.8 Hz, 5.4 Hz, 1H, H^9), 4.21 (dd, 7.4 Hz, 4.3 Hz, 1H, H^2), 3.44 (m, 2H, H^4), 2.81 (dd, 16.4 Hz, 5.1 Hz, 1H, H^{10}), 2.78 (dd, 16.4 Hz, 9.4 Hz, 1H, $H^{10'}$), 2.01 (m, 1H, H^3), 1.89 (m, 1H, $H^{3'}$). ^{13}C NMR (201 MHz, D_2O): δ 176.3 (C1, C8, or C11), 175.8 (C1, C8, or C11), 175.5 (C1, C8, or C11), 160.7 (br, C6 or C7), 159.5 (br, C6 or C7), 157.3 (C5), 69.1 (C2), 50.6

(C9), 37.2 (C4), 36.6 (C10), 32.5 (C3). HRMS (m/z): $[M+H]^+$ $C_{11}H_{18}O_6N_7$, theoretical mass: 344.1313, actual mass: 344.1316.

(*R*/*S*)-Ad^{HA}: To 100 mL of DMF, 6.8 g of adenine (50 mmol) and 7.48 mL of DBU (50 mmol) were added. The mixture was heated to 125°C with rapid stirring. Once the cyanuric acid had fully dissolved, 3.9 mL of DL-α-hydroxy-γ-butyrolactone (50 mmol) were added dropwise. The reaction was stirred at 125°C for 15 hours. The solution was then poured onto a watch glass and allowed to evaporate over several days, eventually producing a thick, brown oil. This material was suspended in 50 mL of water, and the suspension was centrifuged. The supernatant was collected and the pellet was washed twice with 15 mL of water. After washing, the supernatant was again collected and added to the previously collected supernatants. The pooled supernatants were reduced in volume to about 20 mL with a rotary evaporator. To this solution, 2.5 mL of formic acid were added, followed by 60 mL of ethanol. Upon the addition of ethanol, a precipitate formed. The mixture was put in a freezer at -20°C overnight. The mixture was filtered and the solid precipitate washed with cold ethanol and dried under vacuum to give crude (R/S)-Ad^{HA}, which was used without further purification.

(*R*/*S*)-Ad^{HA} Isopropyl Ester: 250 mg of the crude Ad^{HA} were added to a 6 dram vial equipped with a stir bar. 10 mL of isopropanol were added, and then, with rapid stirring, 0.4 mL of thionyl chloride were added dropwise. The vial was sealed and this mixture was heated at 100-110°C for 16 hours. The solvent was then evaporated first by blowing a stream of nitrogen onto the solution, and then by removing any residual solvent under vacuum, giving (*S*)-Ad^{HA} isopropyl ester hydrochloride salt as a white solid in near-

quantitative conversion, as determined by ^{1}H NMR. The solid was used without further purification.

(*S*,*S*)-Ad^{HA}-D: To a 6 dram vial containing approximately 316 mg of Ad^{HA} isopropyl ester hydrochloride (1 mmol), 160 mg of vacuum-dried L-aspartic acid (i.e., (*S*)-D, 1.2 mmol) and 418 mg of vacuum-dried 1,5,7-triazabicyclo[4.4.0]dec-5-ene (TBD, 3 mmol) were added. The glass vial, equipped with a stir bar, was sealed and stirred at 100°C for 48 hours, forming a clear, brown, highly viscous solution. This mixture was dissolved in 2 mL water with 100 µL formic acid, and the pH of the resulting solution was adjusted to 5 by the further addition of formic acid. The solution was loaded onto a column containing 75 mL of QAE Sephadex A-25 resin equilibrated in ammonium formate buffer, 50 mM, pH 5, and separated using a Teledyne Isco CombiFlash Rf+ purification system with a gradient of ammonium formate pH 5, 50 mM, to ammonium formate, pH 5, 500 mM. The fractions containing product were pooled and reduced to about 5 mL with a rotary evaporator. This solution was transferred to a conical tube and brought to a total volume of 20 mL by the addition of water. 60 mL of ethyl acetate were then added, and the solution was shaken vigorously for several minutes. After standing for about an hour, a bed of acetamide formed at the bottom of the aqueous layer. The organic phase was removed and the aqueous phase was heated to bring all of the acetamide into solution. This solution was separated using a Teledyne Isco CombiFlash Rf+ purification system with a RediSep C18Aq 150g Gold column with a 100% water eluent. The solution was injected onto the column in 4 mL portions. The two diastereomers were easily separated. Fractions containing the (*S*,*S*) diastereomer were pooled and reduced to about 5 mL with a rotary evaporator. This solution was then lyophilized to give the product, free of any counter-ions or buffer salts.

^{1}H NMR (700 MHz, D_2O): δ 4.37 (m, 3H,), 4.10 (dd, 7.9 Hz, 4.3 Hz, 1H,), 2.73 (dd, 16.6 Hz, 6.6 Hz, 1H), 2.69 (dd, 16.6 Hz, 5.3 Hz, 1H), 2.31 (m, 1H, H^3), 2.18 (m, 1H).

6.2.4 Prebiotic Synthesis of Nucleobase-Functionalized Hydroxy Acids

Cy^{HA}: In a 6 dram vial, cyanuric acid (129 mg, 1 mmol) was suspended in 20 mL of water. The mixture was heated and stirred until the cyanuric acid fully dissolved. Then, potassium ferrocyanide trihydrate (422.4 mg, 1 mmol) was added and quickly dissolved with stirring. The vial was placed in an oil bath at 45°C. Acrolein (66.8 µL, 1 mmol) was then added, and the vial was sealed and stirred for 18 hours. The reaction was then extracted with 5 × 100 mL dichloromethane, and the pooled extracts were dried over anhydrous magnesium sulfate. The solvent was then removed under vacuum, leaving a brown residue. This residue was dissolved in 5 mL of HCl 6 M and refluxed for 18 hours. The solution was then lyophilized and redissolved in 1 mL of D_2O for NMR analysis. The presence of compound Cy^{HA} was confirmed by spiking with a conventionally synthesized authentic sample.

Mel^{HA}: In a 3 dram vial, melamine (63 mg, 0.5 mmol) was suspended in 5 mL of water with formic acid (18.8 µL, 0.5 mmol). The mixture was heated and stirred until the melamine fully dissolved. Then, potassium ferrocyanide trihydrate (211.2 mg, 0.5 mmol) was added, and a precipitate was rapidly formed that could not be fully redissolved. The vial was placed in an oil bath at 45°C. Acrolein (66.8 µL, 1 mmol) was then added, and the vial was sealed and stirred for 18 hours. The reaction was then extracted with 5 × 50 mL dichloromethane, and the pooled extracts were dried over anhydrous magnesium sulfate. The solvent was then removed under vacuum, leaving a brown residue. This residue was

dissolved in 5 mL of HCl 6 M and refluxed for 18 hours. The solution was then lyophilized and redissolved in 1 mL of D_2O for NMR analysis. The presence of compound Mel^{HA} could not be confirmed by spiking with a conventionally synthesized authentic sample, as this spiking generated new peaks in the ^1H spectrum.

Ad^{HA}: In a 3 dram vial, adenine (67.6 mg, 0.5 mmol) was suspended in 5 mL of water with formic acid (18.8 µL, 0.5 mmol). The mixture was heated and stirred until the adenine fully dissolved. Then, potassium ferrocyanide trihydrate (211.2 mg, 0.5 mmol) was added and quickly dissolved with stirring. The vial was placed in an oil bath at 45°C. Acrolein (66.8 µL, 1 mmol) was then added, and the vial was sealed and stirred for 18 hours. The reaction was then extracted with 5 × 50 mL dichloromethane, and the pooled extracts were dried over anhydrous magnesium sulfate. The solvent was then removed under vacuum, leaving a brown residue. This residue was dissolved in 5 mL of 6 M HCl and refluxed for 18 hours. The solution was then lyophilized and redissolved in 1 mL of D_2O for NMR analysis. The presence of compound Ad^{HA} was confirmed by spiking with a conventionally synthesized authentic sample.

6.2.5 Oligomerization Reactions

Stock solutions of the heterodimers Cy^{HA}-D, Mel^{HA}-D, and Ad^{HA}-D were prepared at 312.5 mM, and stock solutions of water-soluble heterocyclic compounds were prepared at 2.5 M, 1.25 M, 625 mM, and 312.5 mM. To a PCR tube, 1.6 µL of the heterodimer solution (0.8 µL each in the case of mixed oligomerization reactions) and 0.4 µL of the appropriate stock solution of heterocyclic compound were added to give 4, 2, 1, 0.5, 0.25, or 0 molar equivalents of heterocyclic additive. In the case of 4 eq. of heterocyclic additive,

0.8 μL of 2.5 M solution were added. In the case of 0 eq. heterocyclic additive, 0.4 μL water were added. This gave a final heterodimer concentration of 250 mM (or 125 mM each in the case of mixed oligomerizations, and 208 mM (or 104 mM for mixed oligomerizations) for the 4 eq. reactions). Reactions typically had starting pH values around 2 due to the carboxylic acids of the heterodimers. With the caps open, the PCR tubes were transferred to an oven at 85°C and left to dry and react for a specified amount of time. After this time, the residue was redissolved in 20 μL water and sampled for UV-LC/MS analysis.

6.3 Results and Discussion

6.3.1 Prebiotic Synthesis of Nucleobase-Functionalized α-Hydroxy Acids

As mentioned in Chapter 4 (see § 4.5 and Figure 4.7), the prebiotic synthesis of either nucleobase-functionalized α-hydroxy acids and α-amino acids could have plausibly occurred by conjugate addition (i.e., 1,4-addition) of a nucleobase to acrolein, followed by cyanohydrin formation and hydrolysis, in the case of hydroxy acids, or the Strecker synthesis, in the case of amino acids. The prebiotic formation of acrolein [27], and the ability of various nucleobases to robustly react with acrolein [28], was demonstrated by Cleaves. The ability of these acrolein adducts to act as substrates for the Strecker amino acid synthesis has been noted by Cleaves and Bada [29]. The prebiotic feasibility of conjugate addition reactions of nucleobases with Michael acceptors was recently demonstrated by Rodriguez et al [30]. In their study, various nitrogenous heterocyclic compounds (i.e., noncanonical and canonical nucleobases) were found to form conjugate

addition products with Michael acceptors such as acrolein and cyanoacetylene when spark discharge reactions were performed above aqueous solutions of these nucleobases.

Cyanide has the potential to complicate this proposed prebiotic reaction pathway as it is a strong nucleophile that could conceivably out-compete nucleobases for conjugate addition to acrolein. However, with the appropriate cyanide source, it is conceivable that these acrolein adducts could further react to form hydroxy acids or amino acids. Recently, in a study conducted by Smith et al., iron cyanide complexes were detected in CM chondrites [31]. Iron cyanide complexes have previously been implicated as a slow-releasing reservoir of cyanide on the early Earth. In a prebiotic scenario where nucleobases and iron cyanide complexes were present, the atmospheric formation and deposition of acrolein could potentially form nucleobase-functionalized hydroxy acids (or amino acids) by the fast reaction of nucleobases with acrolein, followed by the slow release of cyanide to form cyanohydrins (or amino nitriles).

Reactions of three plausibly prebiotic nucleobases, cyanuric acid, adenine, and melamine, were attempted with acrolein in an aqueous solution containing ferrocyanide (Figure 6.1A). These bases were chosen as cyanuric acid, with an acceptor-donor-acceptor hydrogen-bonding pattern, can form supramolecular assemblies with derivatives of both melamine (donor-acceptor-donor) [13] and adenine (donor-acceptor) [32]. After allowing the nucleobase to react in solution with acrolein and ferrocyanide for 18 hours, the reaction products were extracted (to avoid the danger of generating HCN in the next step), and the extracts were hydrolyzed with acid. These reaction mixtures were analyzed by ^{1}H NMR and spiked with authentic standards of nucleobase-functionalized hydroxy acids (synthesized by conventional means) to confirm their presence. The reactions of cyanuric

acid and adenine were successful in producing a cyanuric acid-functionalized hydroxy acid (Cy^{HA}, Figure 6.1B) and an adenine-functionalized hydroxy acid (Ad^{HA}, Figure 6.1C), respectively. The melamine-functionalized hydroxy acid (Mel^{HA}, Figure 6.1D), however, was not detected.

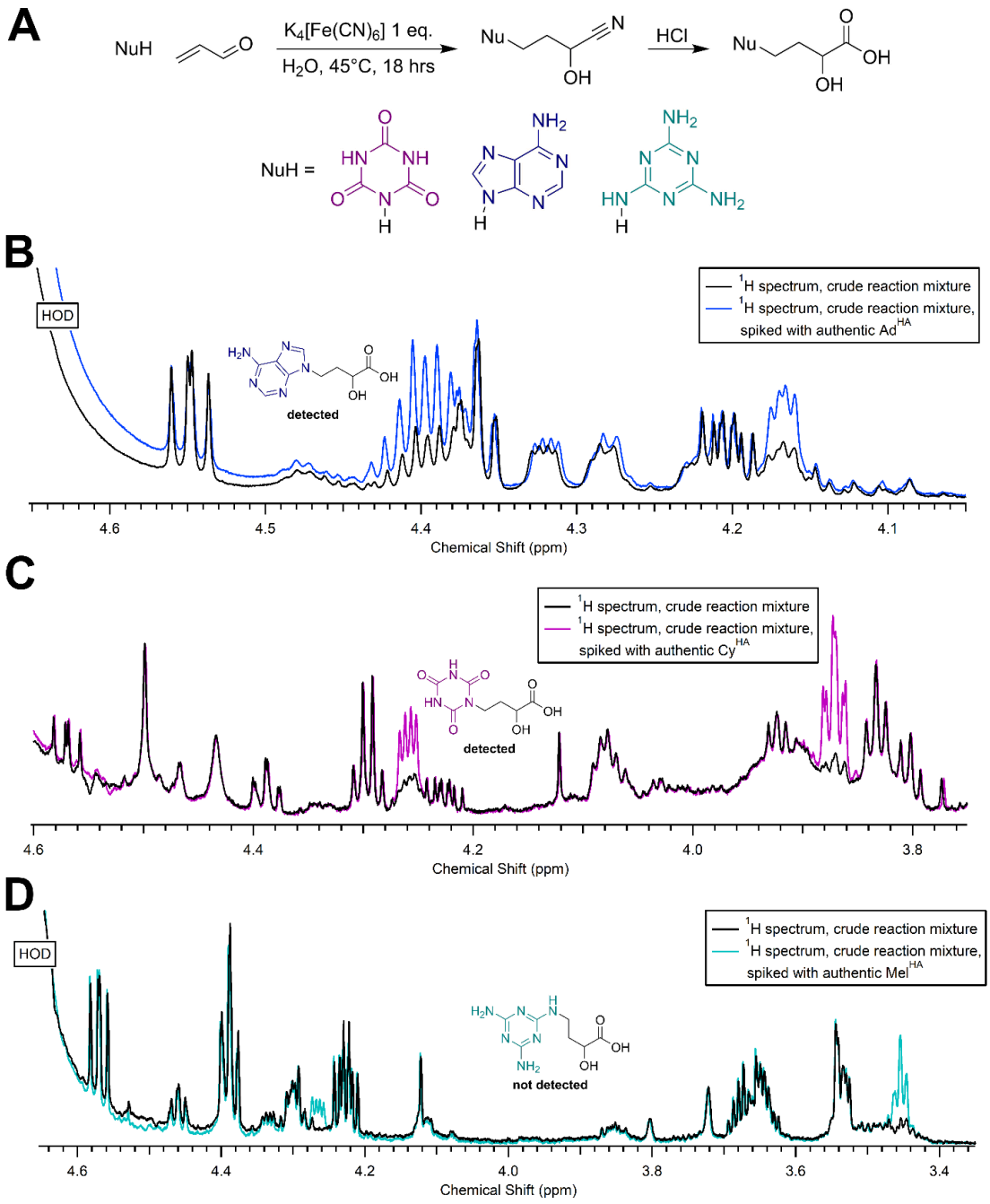


Figure 6.1. Prebiotic synthesis and detection of nucleobase-functionalized amino acids. A. Prebiotic synthesis of nucleobase-functionalized amino acids with acrolein and potassium ferrocyanide. B. ^{1}H spectra of crude adenine reaction mixture after acidic hydrolysis. By spiking with an authentic Ad^{HA} standard, the product was detected in the prebiotic reaction. C. ^{1}H spectra of crude cyanuric acid reaction mixture after acidic hydrolysis. By spiking with an authentic Cy^{HA} standard, the product was detected in the prebiotic reaction. D. ^{1}H spectra of crude melamine reaction mixture after acidic hydrolysis. By spiking with an authentic Mel^{HA} standard, the product was not detected in the prebiotic reaction.

6.3.2 Oligomerization of Model Proto-Nucleic Acid Monomers

Because amino acids and hydroxy acids are produced in the same model prebiotic reactions [33], and because hydroxy acid-*N*-terminated depsipeptide oligomers are formed from dry-down reactions of amino acids and hydroxy acids [34], a model proto-nucleic acid monomer was considered which was composed of a nucleobase-functionalized hydroxy acid linked by an amide bond to an amino acid (Figure 6.2). The compounds (*S*,*S*)-Cy^{HA}-D (a cyanuric acid-functionalized hydroxy acid-amino acid heterodimer, Figure 6.2A), (*S*,*S*)-Ad^{HA}-D (an adenine-functionalized hydroxy acid-amino acid heterodimer, Figure 6.2B), and (*S*,*S*)-Mel^{HA}-D (a melamine-functionalized hydroxy acid-amino acid heterodimer, Figure 6.2C) were synthesized by conventional means and isolated in their free acid forms. (Although Mel^{HA} was not successfully synthesized prebiotically, the melamine-functionalized heterodimer was nevertheless synthesized to further explore oligomerization and self-assembly processes.) As previously alluded to, a cyanuric acid-melamine/adenine pairing system was chosen as derivatives of these heterocycles are well-known to efficiently self-assemble in water [13, 32].

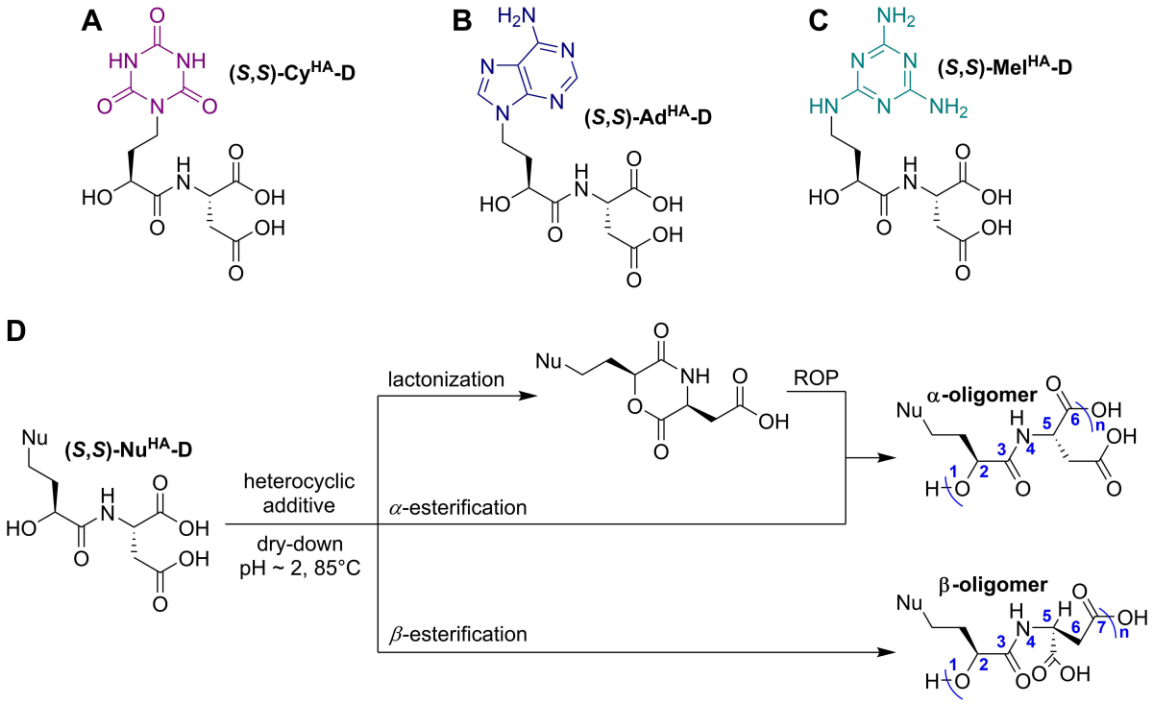


Figure 6.2. The depsipeptide nucleic acid monomers and their prebiotic oligomerization. A. Cy^{HA}-D, the cyanuric acid-functionalized monomer. B. Ad^{HA}-D, the adenine-functionalized monomer. C. Mel^{HA}-D, the melamine-functionalized monomer. D. Prebiotic oligomerization of model monomers. When dried from an acidic aqueous solution containing a heterocyclic additive, oligomerization by esterification occurs. Two backbone linkages are possible: the α-linkage, formed by ring-opening polymerization of the morpholine-2,5-dione form of the monomer, or by direct α-esterification, and the β-linkage, formed by direct β-esterification.

Compounds Cy^{HA}-D, Ad^{HA}-D, Mel^{HA}-D and are expected to oligomerize by esterification to form depsipeptides when dried from an aqueous solution at mildly acidic pH and elevated temperature (Figure 6.2D). In this esterification reaction, lactonization is also expected to occur to form morpholine-2,5-diones, which may act as electrophiles for an in situ ring-opening polymerization reaction, which would ultimately favor oligoesterification at the α-carboxylic acid of the aspartic acid (D) residue to give a backbone with a 6-atom repeat. Direct α-esterification would give the same result. The backbone that would result from this mode of oligomerization would be geometrically similar to previously studied oligopeptide nucleic acid systems that were found to support

base pairing [22]. However, the competing β-esterification may also occur, which would give a 7-atom segment in the backbone of the nascent informational oligomer. It should be noted that α-esters and β-esters may be found in the same oligomer, in a linear or branched manner.

Given the high melting points of the model monomers (>125°C) relative to the temperature of the prebiotic oligomerization reaction (85°C), we anticipated that condensation oligomerization from a dry-down reaction would be inefficient. It was reasoned that the addition of a small molecule compound that would depress the melting point of the monomers would increase the efficiency of oligomerization. Therefore, a suite of heterocyclic additives were surveyed in the oligomerization reactions in varying molar equivalents, with the rationale that these compounds are similar to the soluble components of the ubiquitous polyaromatic organic material present in, for example, carbonaceous chondrites [35]. After drying from aqueous solutions and reacting at 85°C for one week, it was found that 1 equivalent of 2-hydroxypyridine was the most effective additive at enhancing the extent of oligomerization. As expected, it was found that longer reaction times gave greater extents of oligomerization, with the greatest conversion of monomer observed after 8 weeks. The UV-LC/MS analyses of these oligomerization reactions are shown in Figure 6.3. Importantly, when Cy^{HA}-D was combined with either Ad^{HA}-D or Mel^{HA}-D in a single oligomerization reaction, co-oligomers could be detected.

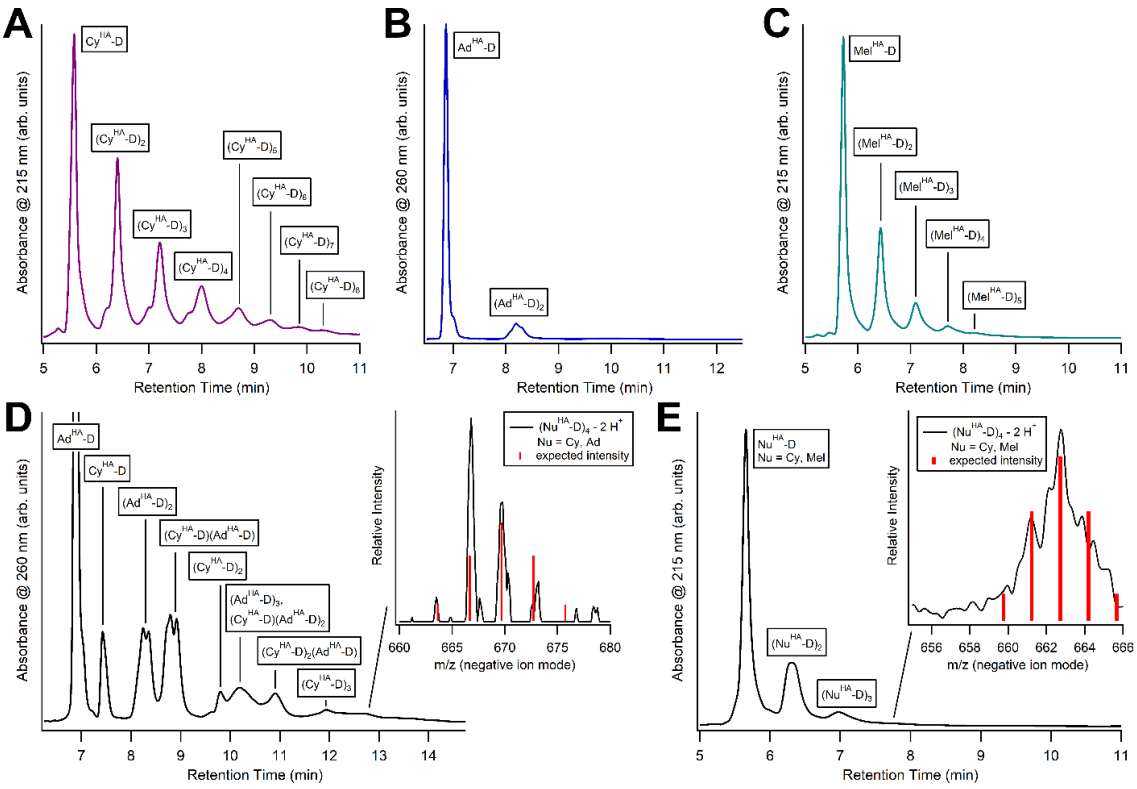


Figure 6.3. UV-LC/MS analysis of the oligomerization reactions of Cy^{HA}-D, Ad^{HA}-D, and Mel^{HA}-D at 85°C with 1 equivalent of 2-hydroxypyridine after 8 weeks. Species are labeled according to the masses identified. A. UV chromatogram of Cy^{HA}-D oligomerization. B. UV chromatogram of Ad^{HA}-D oligomerization. C. UV chromatogram of Mel^{HA}-D oligomerization. D. UV chromatogram of mixed oligomerization of 1:1 Cy^{HA}-D:Ad^{HA}-D. Oligomers of mixed sequence up to 4 units in length are detectable (inset). E. UV chromatogram of mixed oligomerization of 1:1 Cy^{HA}-D:Mel^{HA}-D. Oligomers of mixed sequence up to 4 units in length are detectable (inset).

6.3.3 Self-Assembly and Hydrolytic Stability of Depsipeptide Nucleic Acids

With the confirmation of the ability of these monomers to oligoesterify in a plausibly prebiotic manner, we next turned to molecular self-assembly studies to determine whether or not the monomers or oligomers could form pairing systems in water. Derivatives of cyanuric acid and adenine or melamine are known to form supramolecular assemblies in water by hexad formation featuring three cyanuric acid units and three adenine/melamine units. These planar hexads stack in water to form micron-length

supramolecular fibers, given that the substituents of the assembling units permit this geometry (Figure 6.4A) [9, 13]. Using a ^{1}H NMR-based technique that relies on the loss of integrated signal intensity for assembled molecules [36, 37], it was determined that Mel^{HA}-D formed large assemblies with cyanuric acid in D_2O at 25°C, pD 7.5 with magnesium chloride 25 mM with a minimum assembly concentration of 16 mM. Under these conditions, compound Cy^{HA}-D did not form soluble assemblies with melamine, and Cy^{HA}-D and Mel^{HA}-D could not associate with each other.

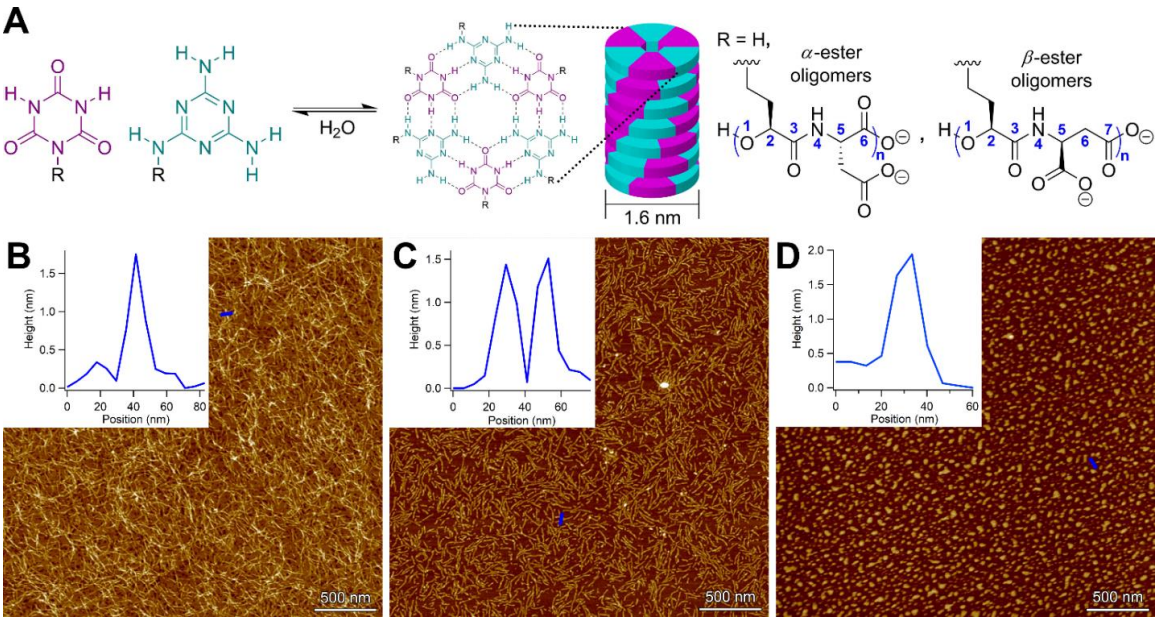


Figure 6.4. Supramolecular assembly of the oligomers of Cy^{HA}-D and Mel^{HA}-D. All AFM images are shown on a 3 µm scale. A. Cyanuric acid- and melamine-derived assembling units associate in water to form hexads which stack to form supramolecular fibers. B. AFM topographical image of Cy^{HA}-D oligomers paired with melamine, pD 6.5. C. AFM topographical image of Mel^{HA}-D oligomers paired with cyanuric acid, pD 6.5. D. AFM topographical image of Cy^{HA}-D/Mel^{HA}-D mixed oligomers with no externally added underivatized pairing heterocycle, pD 7.5, $MgCl_2$ 25 mM.

In order to determine whether or not the oligomers of Cy^{HA}-D and Mel^{HA}-D could self-assemble in water, we performed AFM analyses of crude oligomerization reactions of Cy^{HA}-D and Mel^{HA}-D. Solutions of the oligomers of Cy^{HA}-D or Mel^{HA}-D were prepared at

pD 6.5 and 50 mM (in monomeric equivalents), which were previously determined to not support self-assembly of the monomeric species. Therefore, if assemblies were formed, they would require the incorporation of oligomers. Species formed from the oligomerization of Cy^{HA}-D with a 2-hydroxypyridine additive (after 2 weeks at 85°C) could in-deed form linear assemblies with melamine (Figure 6.4B). Similarly, species formed from the oligomerization of Mel^{HA}-D with 2-hydroxypyridine could also form linear assemblies with cyanuric acid (Figure 6.4C). However, species formed from the co-oligomerization of Cy^{HA}-D and Mel^{HA}-D with 2-hydroxypyridine could not form linear assemblies under these conditions. Upon changing the buffer system to a pD of 7.5 with MgCl2 25 mM, the species formed from the co-oligomerization of Cy^{HA}-D and Mel^{HA}-D could form short assemblies visible by AFM. (Figure 6.4D). These results are not surprising, as the large steric and charge repulsion of the backbones of these oligomers probably prevents association into long fibers [9], but screening of the peripheral negative charge of the oligomers by magnesium facilitates their association into short aggregates, either by pairing of complementary oligomers, or by pairing of oligomers with unreacted monomers.

Because the oligomerization of Cy^{HA}-D and Mel^{HA}-D can result in either α- or β-esters, which may have different propensities for self-assembly, or resistance to hydrolytic degradation, we next endeavored to isolate and characterize individual oligomeric species. Since the oligomerization of Cy^{HA}-D proceeded to the greatest extent, the oligomers formed from this reaction were targeted for isolation. The two major dimeric species (the α-ester, α-$(Cy^{HA}\text{-D})_2$, and the β-ester, β-$(Cy^{HA}\text{-D})_2$) could be purified from the oligomerization reaction of Cy^{HA}-D. These two dimers were isolated in a 2:1 ratio favoring

the α-(Cy^{HA}-D)$_2$, suggesting that in situ ring-opening oligomerization of the initially formed morpholine-2,5-dione was indeed operative in this system. We hypothesized based on structural grounds that α-(Cy^{HA}-D)$_2$ would assemble more efficiently than β-(Cy^{HA}-D)$_2$. The elongated (i.e., all dihedral angles in the anti or s-trans conformation), α-linked depsipeptide backbone of α-(Cy^{HA}-D)$_2$ would place both cyanuric acid residues on the same side of the backbone, and both carboxylate residues on the same side of the backbone, opposite of the cyanuric acid residues (Figure 6.5A), which was expected to facilitate self-assembly. Conversely, the β-linked depsipeptide backbone of β-(Cy^{HA}-D)$_2$ require a near-eclipsed conformation of C_α-C_β bond of the first aspartic acid residue to orient both cyanuric acid residues on the same side of the backbone (Figure 6.5B), which was expected to complicate self-assembly. Surprisingly, it was found that both isomers assembled efficiently with melamine, with a minimum assembly concentration of about 3 mM for both species at a pD of 6.5 with no divalent metal cation. AFM analysis confirmed this finding, with both species producing long supramolecular fibers when paired with melamine at pD 6.5.

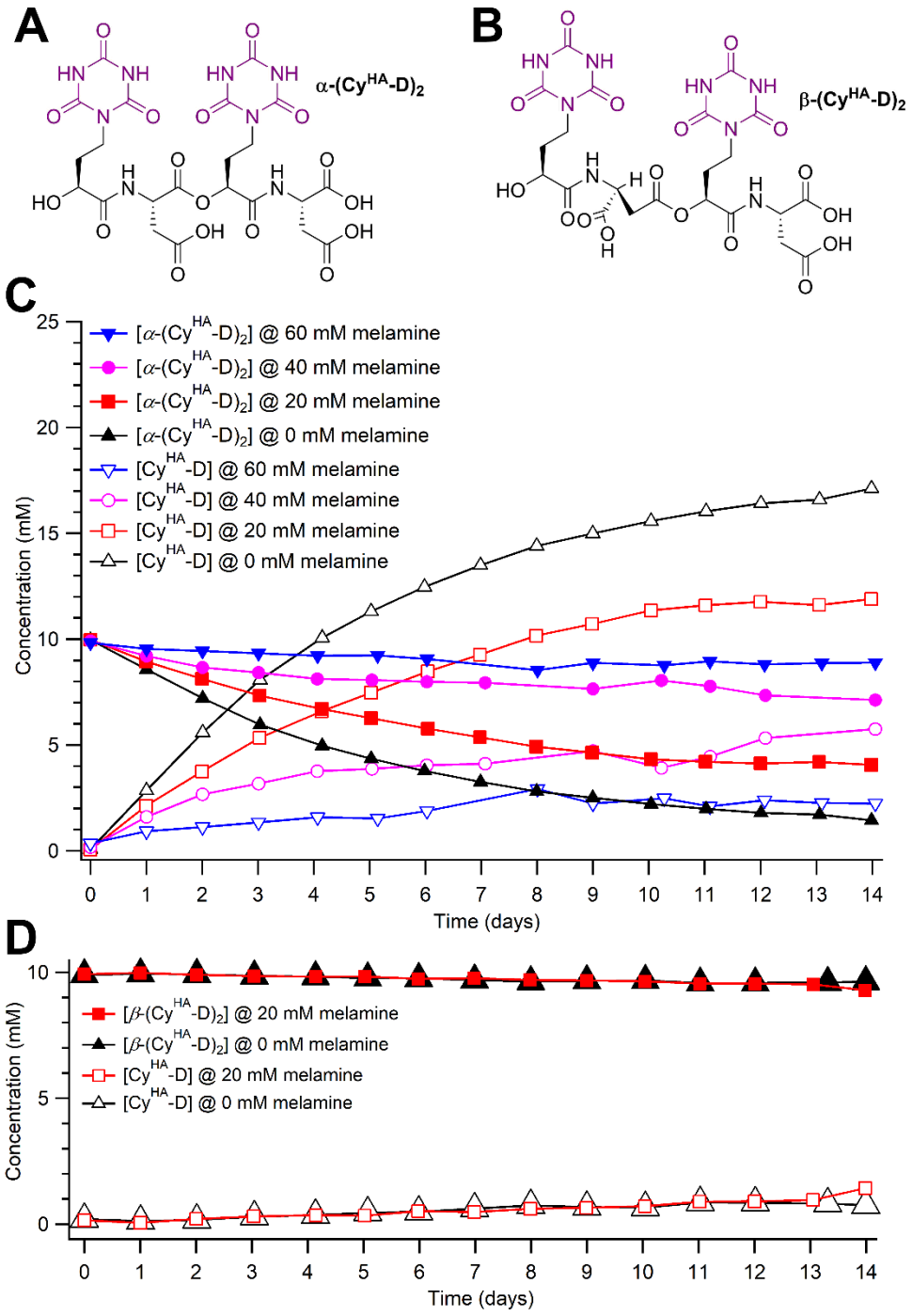


Figure 6.5. Hydrolysis of $(Cy^{HA}\text{-}D)_2$ at pD 6.5 and its retardation by supramolecular assembly with melamine. All reported concentrations are normalized according to the formula $[(Cy^{HA}\text{-}D)_2] + ½[Cy^{HA}\text{-}D] = 10$ mM. A. Compound α-$(Cy^{HA}\text{-}D)_2$, the α-ester dimer of Cy^{HA}-D. B. Compound β-$(Cy^{HA}\text{-}D)_2$, the β-ester dimer of Cy^{HA}-D. C. Hydrolysis of α-$(Cy^{HA}\text{-}D)_2$ with varying concentrations of melamine. D. Hydrolysis of β-$(Cy^{HA}\text{-}D)_2$ with or without melamine at 20 mM.

Anticipating the reasonable objection that the ester would be too hydrolytically unstable of a chemical linkage to support a hereditary polymer system, we next performed hydrolysis studies using the isolated dimers of Cy^{HA}-D. Com-pounds α-(Cy^{HA}-D)$_2$ and β-(Cy^{HA}-D)$_2$ were incubated at 10 mM at 25°C in D_2O at pD 6.5 with or without melamine, forming hydrogels when melamine was present. Over a period of 14 days, the concentrations of α-(Cy^{HA}-D)$_2$ or β-(Cy^{HA}-D)$_2$ and monomeric Cy^{HA}-D were measured by UV-LC/MS (Figure 6.5). It was found that the rate of hydrolysis of α-(Cy^{HA}-D)$_2$ over this period of time was retarded by increasing concentrations of melamine (Figure 6.5C), suggesting that α-(Cy^{HA}-D)$_2$, in the assembled state, is resistant to hydrolysis. Interestingly, the rate of hydrolysis of β-(Cy^{HA}-D)$_2$ was much slower than that of α-(Cy^{HA}-D)$_2$, and any effect of melamine, if existent, was not appreciable over the timespan of the experiment (Figure 6.5D). The faster rate of hydrolysis of α-(Cy^{HA}-D)$_2$ compared to β-(Cy^{HA}-D)$_2$ could be due to a backbiting mechanism of hydrolysis available to α-(Cy^{HA}-D)$_2$ in which the terminal hydroxyl group, upon transient deprotonation, performs intramolecular attack on the ester moiety to form a 6-membered ring intermediate, which ultimately leaves as the morpholine-2,5-dione of Cy^{HA}-D, along with one equivalent of parent Cy^{HA}-D. The morpholine-2,5-dione would then be rapidly hydrolyzed by the solvent, preventing its detection. This mechanism of hydrolysis would not be possible for β-(Cy^{HA}-D)$_2$, as intramolecular attack of the ester by the terminal hydroxyl group would result in a 7-membered ring intermediate [38].

6.4 Conclusions

The results presented here demonstrate that proto-nucleic acid oligoesters are easily generated under plausibly prebiotic conditions from plausible monomers. The nucleobase-functionalized hydroxy acids Cy^{HA} and Ad^{HA} can be formed in a plausibly prebiotic manner. Compounds Cy^{HA}-D, Ad^{HA}-D, and Mel^{HA}-D oligoesterify when dried from an aqueous solution at 85°C at a mildly acidic pH. The acidic pH is derived from the acidic moieties on the monomers themselves, and does not require the external addition of acid. The use of heterocyclic compounds to enhance the extent of the oligomerization reaction is, in fact, more reflective of a true prebiotic scenario, in which the proto-nucleic acid monomers would probably not have been present in pure form, but would have been contaminated with extraneous carbonaceous matter. The free energy required to produce the oligoesters is derived from a temporal thermal gradient, as would have been common on the prebiotic Earth. No activation chemistry is required. This stands in contrast to many prebiotic syntheses of RNA oligomers, which require pre-activated mononucleotides [39] or the additive of a chemical activating agent of questionable prebiotic viability [40]. Furthermore, compounds Cy^{HA}-D and Mel^{HA}-D, and their oligomers, possess the propensity for self-assembly in water. Most germanely, the species formed from the co-oligomerization of Cy^{HA}-D and Mel^{HA}-D have the ability to self-assemble, as would be required for information transfer in a proto-genetic system.

The primary objection to the viability of an ester-linked proto-nucleic acid is perhaps the known hydrolytic instability of esters. If, however, a candidate polymer has the ability to function as a genetic molecule, it must necessarily undergo molecular recognition, (i.e., intermolecular self-assembly). Once in an assembled structure, the

oligoesters sample fewer conformational states, and are therefore less likely to distort to a structure close to the transition state of hydrolysis. This may provide a means of selection of certain sequences over others based on their propensities for self-assembly. Furthermore, a proto-genetic system which forms totally irreversible chemical linkages is not necessarily desirable for chemical evolution, as monomers would be rapidly exhausted, preventing further evolution [41]. The system described here forms oligomers that are indeed stabilized against hydrolysis by self-assembly, but can also selectively hydrolyze.

Finally, although the system described here is a model system, it demonstrates certain key principles that may also apply to other prebiotic systems. Namely, the ester, by virtue of facile prebiotic formation, support of oligomer self-assembly, and resistance to hydrolysis in an assembled state, is a viable chemical linkage for the earliest informational molecules that may have arisen on the early Earth.

6.5 References

[1] Krishnamurthy, R., Experimentally investigating the origin of DNA/RNA on early Earth. Nature Communications, 2018. 9(1): p. 5175.

[2] Joyce, G.F. and L.E. Orgel, Prospects for understanding the origin of the RNA world, in The RNA World, Second Edition: The Nature of Modern RNA Suggests a Prebiotic RNA World, R.F. Gesteland and J.F. Atkins, Editors. 1999, Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY. p. 49-77.

[3] Hud, N.V., Searching for lost nucleotides of the pre-RNA World with a self-refining model of early Earth. Nature Communications, 2018. 9(1): p. 5171.

[4] Hud, N.V., et al., The origin of RNA and 'My Grandfather's Axe'. Chem. Biol., 2013. 20: p. 466-474.

[5] Kolb, V.M., J.P. Dworkin, and S.L. Miller, Alternative bases in the RNA world: The prebiotic synthesis of urazole and its ribosides. J. Mol. Evol., 1994. 38: p. 549-557.

[6] Bean, H.D., et al., Formation of a β-pyrimidine nucleoside by a free pyrimidine base and ribose in a plausible prebiotic reaction. J. Am. Chem. Soc., 2007. 129: p. 9556-9557

[7] Chen, M.C., et al., Spontaneous prebiotic formation of a β-ribofuranoside that self-assembles with a complementary heterocycle. Journal of the American Chemical Society, 2014. 136(15): p. 5640-5646.

[8] Kim, H.J. and S.A. Benner, Prebiotic glycosylation of uracil with electron-donating substituents. Astrobiology, 2015. 15(4): p. 301-6.

[9] Cafferty, B.J., et al., Spontaneous formation and base pairing of plausible prebiotic nucleotides in water. Nature Communications, 2016. 7: p. 11328.

[10] Becker, S., et al., Wet-dry cycles enable the parallel origin of canonical and non-canonical nucleosides by continuous synthesis. Nature Communications, 2018. 9(1): p. 163.

[11] Gonzalez, M.A., et al., Facile preparation of C-glycosylbarbiturates and C-glycosylbarbituric acids. Carbohydrate Research, 1986. 158: p. 53-66.

[12] Fialho, D.M., et al., Glycosylation of a model proto-RNA nucleobase with non-ribose sugars: implications for the prebiotic synthesis of nucleosides. Organic & Biomolecular Chemistry, 2018. 16: p. 1263-1271.

[13] Ma, M. and D. Bong, Determinants of cyanuric acid and melamine assembly in water. Langmuir, 2011. 27(14): p. 8841-53.

[14] Eschenmoser, A., Chemical etiology of nucleic acid structure. Science, 1999. 284(5423): p. 2118-2124.

[15] Pinheiro, V.B., et al., Synthetic genetic polymers capable of heredity and evolution. Science, 2012. 336(6079): p. 341-344.

[16] Lohrmann, R. and L.E. Orgel, Urea-inorganic phosphate mixtures as prebiotic phosphorylating agents. Science, 1971. 171(3970): p. 490-494.

[17] Burcar, B., et al., Darwin's Warm Little Pond: A one-pot reaction for prebiotic phosphorylation and the mobilization of phosphate from minerals in a urea-based solvent. Angew. Chem. Int. Ed. Engl., 2016. 55: p. 13249-13253.

[18] Gibard, C., et al., Phosphorylation, oligomerization and self-assembly in water under potential prebiotic conditions. Nat. Chem., 2018. 10(2): p. 212-217.

[19] Keefe, A.D. and S.L. Miller, Are polyphosphates or phosphate esters prebiotic reagents? J. Mol. Evol., 1995. 41: p. 693-702.

[20] Bean, H.D., et al., Glyoxylate as a backbone linkage for a prebiotic ancestor of RNA. Origins of Life and Evolution of the Biosphere, 2006. 36(1): p. 39-63.

[21] Nelson, K.E., M. Levy, and S.L. Miller, Peptide nucleic acids rather than RNA may have been the first genetic molecule. Proceedings of the National Academy of Sciences of the United States of America, 2000. 97(8): p. 3868-3871.

[22] Mittapalli, G.K., et al., Mapping the landscape of potentially primordial informational oligomers: oligodipeptides and oligodipeptoids tagged with triazines as recognition elements. Angew. Chem., Int. Ed. Engl., 2007. 46(14): p. 2470-2477.

[23] De Graaf, R.M. and A.W. Schwartz, Thermal synthesis of nucleoside H-phosphonates under mild conditions. Origins of Life and Evolution of the Biosphere, 2005. 35(1): p. 1-10.

[24] Lonnberg, T., Sulfurization of H-Phosphonate Diesters by Elemental Sulfur under Aqueous Conditions. ACS Omega, 2017. 2(8): p. 5122-5127.

[25] Harada, K. and L.E. Orgel, Template-Directed Oligomerization of 5'-Deoxy-5'-Nucleosideacetic Acid-Derivatives. Origins of Life and Evolution of the Biosphere, 1990. 20(2): p. 151-160.

[26] Mamajanov, I., et al., Ester formation and hydrolysis during wet-dry cycles: generation of far-from-equilibrium polymers in a model prebiotic reaction. Macromolecules, 2014. 47: p. 1334-1343.

[27] Cleaves, H.J., The Prebiotic Synthesis of Acrolein. Monatshefte für Chemie, 2003. 134(4): p. 585-593.

[28] Cleaves, H.J., The Reactions of Nitrogen Heterocycles with Acrolein: Scope and Prebiotic Significance. Astrobiology, 2002. 2(4): p. 403-415.

[29] Cleaves, H.J. and J.L. Bada, The Prebiotic Chemistry of Alternative Nucleic Acids, in Genesis - In The Beginning: Precursors of Life, Chemical Models and Early Biological Evolution, J. Seckbach, Editor. 2012, Springer Science+Business Media Dordrecht. p. 3-33.

[30] Rodriguez, L.E., et al., Nitrogen heterocycles form peptide nucleic acid precursors in complex prebiotic mixtures. Sci. Rep., 2019. 9(1): p. 9281.

[31] Smith, K.E., et al., Organometallic compounds as carriers of extraterrestrial cyanide in primitive meteorites. Nat. Commun., 2019. 10(1): p. 2777.

[32] Li, C., et al., Formation of supramolecular assemblies and liquid crystals by purine nucleobases and cyanuric acid in water: implications for the possible origins of RNA. Physical Chemistry Chemical Physics, 2016. 18(30): p. 20091-20096.

[33] Miller, S.L. and H.C. Urey, Organic compound synthesis on the primitive Earth. Science, 1959. 130(3370): p. 245-251.

[34] Forsythe, J.G., et al., Ester-Mediated Amide Bond Formation Driven by Wet-Dry Cycles: A Possible Path to Polypeptides on the Prebiotic Earth. Angewandte Chemie-International Edition, 2015. 54(34): p. 9871-9875.

[35] Sephton, M.A., Organic compounds in carbonaceous meteorites. Natural Product Reports, 2002. 19(3): p. 292-311.

[36] Cafferty, B.J., et al., Efficient self-assembly in water of long noncovalent polymers by nucleobase analogues. Journal of the American Chemical Society, 2013. 135(7): p. 2447-2450.

[37] Cafferty, B.J., et al., Ultra-sensitive pH control of supramolecular polymers and hydrogels: pK(a) matching of biomimetic monomers. Chemical Science, 2014. 5(12): p. 4681-4686.

[38] de Jong, S.J., et al., New insights into the hydrolytic degradation of poly(lactic acid): participation of the alcohol terminus. Polymer, 2001. **42**: p. 2795-2802.

[39] Ferris, J.P., et al., Synthesis of long prebiotic oligomers on mineral surfaces. Nature, 1996. 381(6577): p. 59-61.

[40] Jauker, M., H. Griesser, and C. Richert, Copying of RNA Sequences without Pre-Activation. Angewandte Chemie-International Edition, 2015. 54(48): p. 14559-14563.

[41] Walker, S.I., M.A. Grover, and N.V. Hud, Universal sequence replication, reversible polymerization and early functional biopolymers: A model for the initiation of prebiotic sequence evolution. PLoS ONE, 2012. 7(4): p. e34166.

CHAPTER 7. CONCLUSION

7.1 Recapitulation

The nucleic acids, RNA and DNA, are the vectors of heredity in extant life. Their central role in cellular processes is self-evident from the Central Dogma [1]. Because of this central role, the study of the origin of the nucleic acids has been a major thread in origins-of-life research [2, 3]. The RNA world hypothesis highlights the importance of RNA to early forms of life, and because of this, some have inferred that RNA was the first polymer of life to spontaneously arise from prebiotic chemical processes on the early Earth [3]. Although some interesting syntheses of nucleosides [4], nucleotides [5], and short RNA oligomers [6] have been proposed, it is worthwhile to consider an alternative hypothesis in which RNA is the product of evolution [7].

It is important to note that, although RNA may be the oldest extant biopolymer, it does not follow that it was necessarily the *first* biopolymer. DNA has been considered an evolutionary descendent of RNA (although some evidence to the contrary has been presented) by modification of the backbone sugar (ribose to 2-deoxyribose) and of one nucleobase (uracil to thymine). In the same way, RNA may be the descendent of some now extinct nucleic acid. In fact, there may have been an evolutionary series of pre-RNA genetic polymers that began with some primordial genetic polymer, the proto-nucleic acid. This proto-nucleic acid was not optimized in its structure like RNA and DNA, but was more easily formed under plausible early-Earth (i.e., prebiotic) conditions [7].

7.2 Physical Organic Principles Governing the Formation of Proto-Nucleic Acids

7.2.1 Physical Organic Principles Governing Proto-Nucleic Acid Nucleobase Selection

Because nitrogenous heterocyclic compounds, including, but not limited to, the canonical nucleobases, form complementary hydrogen-bonded (Watson-Crick-like) associations in water [8, 9], they are an obvious choice for a mechanism of transduction of information in a chemical system. Although other mechanisms of information transfer that do not rely on complementary hydrogen bonding have been suggested [10], the ubiquity of nitrogenous heterocyclic compounds in model prebiotic experiments [11] and in meteorites [12] makes them strong candidates for recognition units in a proto-nucleic acid. Furthermore, if another chemical mechanism for information transduction and replication had been present at the origin of life, it is not clear how such a system would transform into the Watson-Crick base-pairing system of extant nucleic acids.

The canonical nucleobases of RNA are, according to certain criteria, optimal for the function of RNA [9]. However, this does not necessarily mean that the canonical bases were present throughout the entire evolutionary lineage of nucleic acids. This first became evident when it was found that the canonical nucleobases do not spontaneously form nucleosides with ribose in water [13, 14].

As explained in Chapters 2, 3, and 4, most of the prebiotically relevant nitrogenous heterocyclic compounds are nucleophilic in their chemical reactivity. However, the strengths and modes of nucleophilicity vary greatly. For the canonical nucleobases, the sites of canonical glycosidic bond formation are all *monovalent* nucleophilic sites; i.e., in any reaction product or intermediate, the maximum possible bond order that can be formed

is 1. For nucleosidation reactions with an unactivated sugar (such as ribose), the mechanism usually invoked [15] requires nucleophiles to be at least *divalent*. From this principle alone, it is obvious that the canonical nucleobases, no matter how great their absolute nucleophilic strength, cannot successfully react with ribose in water to form nucleosides.

In extant nucleotide salvage pathways, which form nucleotides by direct ribosylation of nucleobases, this problem is overcome by activating ribose at position 1 (the anomeric position) with a good leaving group in the α-orientation. For example, in purine salvage, 5-phosphoribose-1-pyrophosphate (PRPP) is the active electrophile. A purine nucleobase, such as adenine, guanine, or hypoxanthine, acts as a nucleophile at *N*7 and attacks PRPP at the anomeric position in the β-orientation with simultaneous departure of the pyrophosphate leaving group. Put more tersely, the reaction proceeds by an S_N2 mechanism, and S_N2 reactions emphatically only require a monovalent nucleophile.

Aqueous glycosylation reactions can proceed without activation with a sufficiently strong divalent nucleophile. For this reason, adenine can react with ribose at its exocyclic amino group at position 6 to form an atypical adenine nucleoside [13]. Similarly, the noncanonical nucleobases 2,4,6-triaminopyrimidine (TAP) [16], barbituric acid, and melamine [17] can react with ribose at their various divalent nucleophilic sites to form noncanonical nucleosides. This reactivity is not limited to ribose; many other sugars successfully react with TAP [18] and barbituric acid [19, 20] to form glycosides.

The nucleobases TAP, barbituric acid, melamine, and cyanuric acid are especially interesting in the context of proto-RNA formation as these, and some structurally related nitrogenous heterocycles, are capable of efficient supramolecular self-assembly in water

[17, 21-23]. Because of this, monomers and short oligomers appended with these nucleobases readily self-assemble in water, whereas canonical nucleotide self-assembly does not occur at the monomer level. Because prebiotic oligomerization reactions typically do not produce long oligomers [24], the self-assembly of short, prebiotically generated informational oligomers may have been important for early selection processes (see below).

Although nucleobase glycosylation clearly became important at some early stage in the evolution of life, it was not necessarily the first mechanism of informational monomer generation. As discussed in Chapters 4, 5, and 6, a number of different plausibly prebiotic classes of electrophile are capable of reacting with nucleobases. In particular, 5-aminopyrimidines are capable of reacting with esters and thioesters to form candidate proto-nucleic acid monomers. Furthermore, most nucleophilic nucleobases are capable of reacting with electron-withdrawing group-substituted unsaturated systems (i.e., Michael acceptors) such as acrolein to form γ-nucleobase-functionalized aldehydes. These aldehydes are substrates for α-amino acid formation by the Strecker synthesis, or α-hydroxy acid synthesis via cyanohydrin formation. Because of the robustness of this reaction, these adducts may have been especially important to the formation of proto-nucleic acids. Because so many nucleobases successfully undergo conjugate addition reactions [25, 26], the selection of nucleobases in proto-nucleic acids may have occurred through self-assembly, where oligomers that contained self-assembling nucleobases were more resistant to hydrolysis, and those that contained nucleobases incapable of self-assembly were more readily hydrolyzed.

7.2.2 Physical Organic Principles Governing Proto-Nucleic Acid Backbone Selection

There are two chemical processes that the components of a backbone of an informational polymer must be capable of: nucleosidation (or the process of appending nucleobases to the backbone in general) and polymerization. Candidate backbone components must be electrophilic enough (and with the proper mode of reactivity with regard to valency) to react with nucleobases, and must be able to form oligomers of the proper repeat length and with chemical linkages that are formed in a prebiotically facile manner.

Ribose phosphates are inefficient at ribosylating the canonical nucleobases (except for ribose-1-phosphates [27]) and at polymerizing in a plausibly prebiotic manner. However, ribose, and other sugars, can efficiently glycosylate divalent noncanonical nucleobases such as urazole [15], TAP [16], and barbituric acid [17]. Although a number of different sugars have been shown to form backbones that geometrically permit the formation of duplexes [28, 29], mixed backbones can be deleterious to self-assembly [30], and sugars are typically not produced selectively by prebiotic processes [31]. For these reasons, sugars may have become important at some early stage of the evolution of life, but not necessarily at the origin of life.

As mentioned above, Michael acceptors such as acrolein efficiently react with nucleobases. Although some nucleobases react to form cyclic adducts, many are capable of forming γ-nucleobase-functionalized aldehydes [25], which can be prebiotically converted to α-hydroxy acids and α-amino acids [32]. A proto-nucleic acid monomer system based on amino acids and hydroxy acids is advantageous for several reasons. First,

mixtures of amino acids and hydroxy acids, under wet-dry cycling conditions, are known to form depsipeptide (mixed ester-amide) oligomers [33]. Second, if heterodimeric depsipeptide cassettes, with one residue featuring a nucleobase, and the other featuring a solubilizing moiety, can be formed, the geometry of the backbone formed from oligomerization of this heterodimer permits supramolecular assembly (by either forming a 6-atom or 7-atom backbone repeat; see Chapter 6). Third, this system is dynamic. Although the ester is stabilized enough by supramolecular assembly to permit the persistence of proto-nucleic acid oligomers, it can also be hydrolyzed to recycle monomers [34]. The recycling of monomers is important for evolution [35], as backbones that form irreversible bonds are locked into their initial sequences. Once the reservoir of monomers has been depleted, no further evolution could occur.

Although not explicitly studied here, many questions of backbone assembly still remain. For example, α-hydroxy acids and α-amino acids are chiral, and in a more prebiotically realistic scenario, heterochiral sequences will initially form from dry-down oligomerization reactions. However, since the depsipeptide oligomers feature hydrolyzable esters, and since the rate of hydrolysis of these esters decreases when in a supramolecular assembly, chiral selection may be possible if heterochiral sequences are less competent at self-assembly than homochiral sequences. Wet-dry cycling of such as system may be capable of enriching individual oligomers in one handedness of amino acid or hydroxy acid.

7.3 Theoretical Considerations Governing the Formation and Early Evolution of Proto-Nucleic Acids

7.3.1 Informational Synonymity of Nucleobases in Proto-Nucleic Acids

The ability of nucleic acids to store information in extant life is a consequence of complementary base pairing. In RNA, because guanine and cytosine always specify each other, and because adenine and uracil always specify each other, knowledge of the sequence of a parent nucleic acid strand gives almost complete certainty of the sequence of a daughter nucleic acid strand. In this context, the meaning of *information* is the ability to accurately predict the outcome of a random variable (among many possible outcomes) based on knowledge of the state of another random variable (among many possible states). For nucleic acid replication, the initial random variable is the nucleotide sequence of the parent strand (for which 4^n states are possible, where n equals the length of the parent nucleic acid), and the final random variable is the sequence of the daughter strand (for which 4^n states, or outcomes, are also possible, excluding insertions and deletions). This definition of information makes intuitive sense: if some mode of replication is error-prone, this is equivalent to saying that knowledge of the sequence of the parent strand does not allow an accurate prediction of the sequence of the daughter strand, and little information is successfully transmitted. This informal description of information can easily be formalized by introducing the concept of Shannon entropy, but will not be necessary for the present discussion. Instead, consider a simple example of the information in extant nucleic acids. In DNA and RNA, the information density is equal to 2 bits per base (because four states can be specified by a single nucleotide symbol, which would require two symbols to specify (2^2) in a binary system). In an arbitrary informational system, the

informational density, measured in bits, will similarly be equal to the logarithm (base 2) of the number of states specifiable per recognitions unit.

Consider a prebiotic scenario in which many (more than four) nitrogenous heterocyclic compounds are present that can potentially react with some backbone component and become incorporated into a proto-nucleic acid. This is not necessarily a purely abstract scenario, as many heterocyclic compounds can be produced by a single model prebiotic reaction [11], and that Michael acceptors such as acrolein can react with many nucleobases that can eventually be transformed into nucleobase-functionalized hydroxy acids or amino acids [32]. With such a large repertoire of monomers to choose from, and in the absence or error-correcting mechanisms facilitated by enzymes, it seems unlikely that information will ever be transferred faithfully in any replication event.

Although any given residue of a proto-nucleic acid may specify a variety of other monomers, these monomers will not necessarily be specified with equal probability. The ability of the canonical nucleobases to specify their partners selectively is the result of complementary base pairing. Although many heterocyclic compounds will be present in a hypothetical prebiotic mixture, these different classes many display similarities in their hydrogen bonding patterns. Consider, for example, melamine and TAP. Both display a donor-acceptor-donor (D-A-D) hydrogen bonding pattern, and as a consequence, both have the capacity to pair with cyanuric acid (with an acceptor-donor-acceptor (A-D-A) hydrogen bonding pattern) to form supramolecular assemblies. A similar phenomenon may occur in proto-nucleic acids: with many heterocyclic compounds to choose from, the ability of any given nucleobase, with some specific hydrogen bonding pattern, to discriminate between bases of a complementary hydrogen-bonding pattern will be low, but the ability to

discriminate complementary hydrogen-bonding patterns from non-complementary patterns will be comparably greater. In this way, many nucleobases will be *informationally synonymous*: members of a single class of synonyms will almost always specify members of the complementary class of synonyms. Consider, for example, a system composed of eight bases: melamine, TAP, 2,6-diaminopurine, adenine, cyanuric acid, barbituric acid, urazole, and uracil (Figure 7.1). Although this system would superficially seem to have an information density of 3 bits (= $\log_2 8$), any member of the adenine family (the D-A-D pattern: melamine, TAP, 2,6-diaminopurine, and adenine) can specify any other member of the uracil family (the A-D-A pattern: cyanuric acid, barbituric acid, urazole, and uracil). Since only two specifiable states exist (the D-A-D state and the A-D-A state), the true information density of this system is 1 bit (= $\log_2 2$). In this manner, the existence of many heterocyclic moieties is not totally detrimental to informational storage capacity.

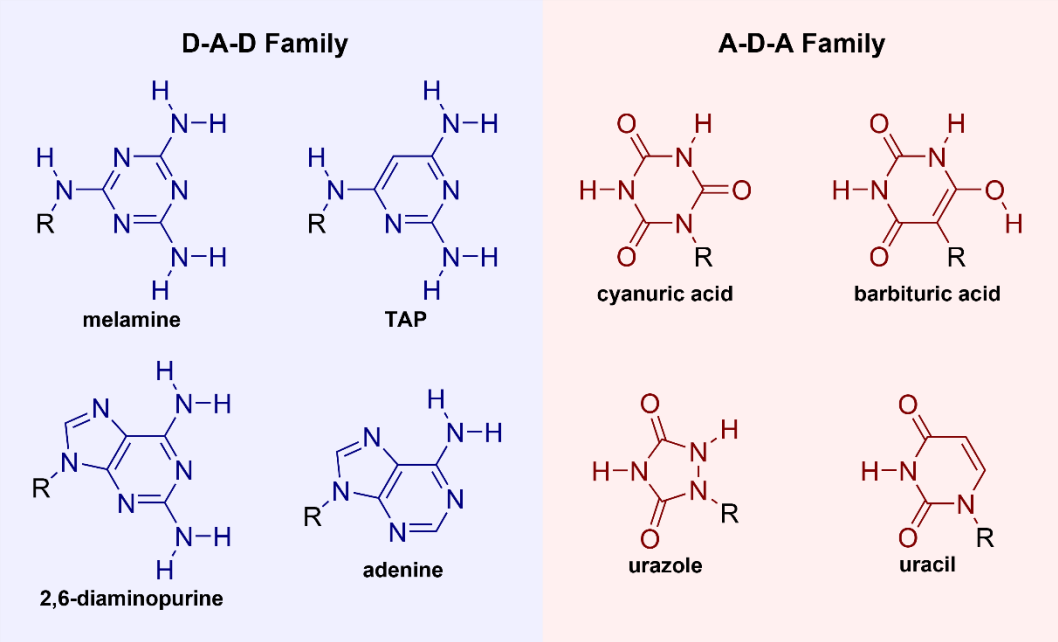


Figure 7.1. Prebiotic heterocyclic compounds partitioned into informationally synonymous classes. Melamine, TAP, 2,6-diaminopurine, and adenine, with a donor-acceptor-donor (D-A-D) hydrogen-bonding pattern, form one family of synonymous symbols. Cyanuric acid, barbituric acid, urazole, and uracil, with an acceptor-donor-acceptor (A-D-A) hydrogen-bonding pattern, form another family of synonymous symbols.

7.3.2 Evolutionary Trajectories of Nucleic Acids

The nature of the evolutionary process that led to RNA (and DNA) is unclear, and a number of scenarios have been offered. To simplify and systematize this discussion, consider two (somewhat artificial) chemical parameters for informational polymers: heterogeneity and absolute chemical identity. Heterogeneity describes how uniform the elements of the information polymer are . For example, RNA and DNA are low-heterogeneity polymers: their backbone components rarely vary, and there are usually only four informational elements that present as two complementary base pairs. RNA can be thought of as marginally more heterogeneous than DNA since nucleobase modifications are common in tRNAs. Absolute chemical identity describes the ability to distinguish one

type of informational polymer from another. RNA and DNA, although similar, are chemically distinct in their backbone sugar, and in the nucleobase that pairs with adenine. In an absolute sense, they are distinct types of polymer. These parameters are not ideal because they are emphatically not orthogonal: a highly heterogeneous polymer does not have a well-defined absolute chemical identity. Nevertheless, they will suffice for the following discussion.

The origin and evolution of nucleic acids can be represented in an abstract 3-dimensional chemical space of heterogeneity, absolute chemical identity, and time. In this chemical space, different models for the origin of RNA can be represented and compared. To further simplify this analysis, this 3-dimensional space can be projected onto a 2-dimensional space of heterogeneity and time, with absolute chemical identity represented with color.

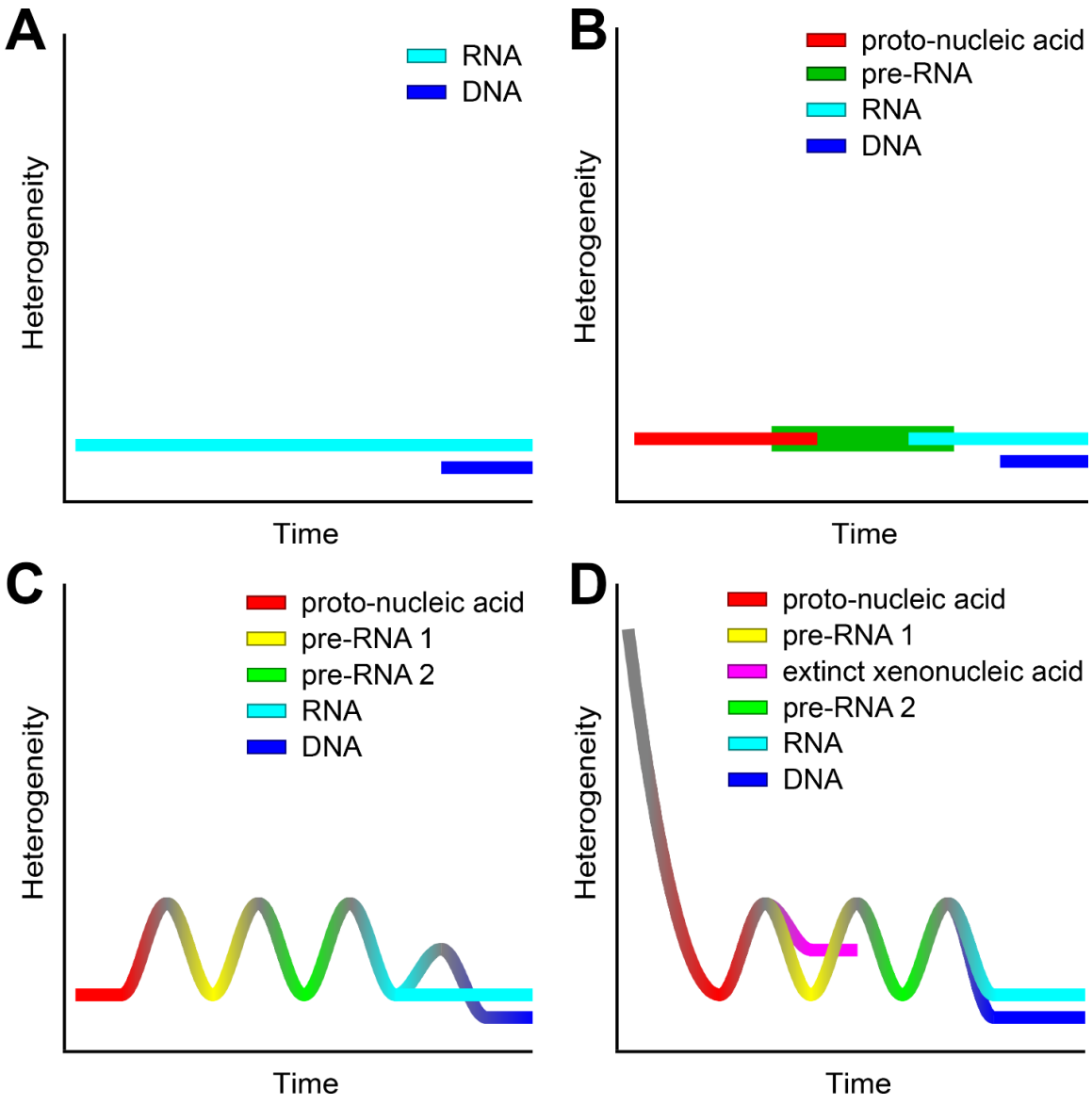


Figure 7.2. Evolutionary trajectories in different models of the emergence of nucleic acids in a chemical space of polymer heterogeneity as a function of time. A. The rudimentary RNA-first model, in which RNA spontaneously emerges from prebiotic chemistry. Sometime later, DNA forms. B. A simplified version of the Cleaves-Bada model of the origin of nucleic acids in which a proto-nucleic acid spontaneously emerges and templates the formation of a descedent pre-RNA, which templates the formation of RNA. C. The Grandfather's Axe model of the origin of nucleic acids in which transitions from one nucleic acid form to another proceed through heterogeneous intermediates. D. An example of a scenario for the origin of nucleic acids in which the proto-nucleic acid emerges from an initially highly heterogeneous polymer and evolves through heterogeneous transitions to eventually give RNA and DNA. The evolution of possible xenonucleic acids, now extinct, are also present.

There are many possible trajectories in this chemical space that could have led to RNA (and DNA). Consider first the traditional RNA-first model, in which RNA is the direct product of prebiotic chemistry, with no predecessors (Figure 7.2A) [36]. In this model, RNA emerges intact and complete, with a well-defined absolute chemical identity and low chemical heterogeneity. Sometime later, DNA spontaneously emerges, perhaps in a template-directed manner from RNA, also with low heterogeneity (perhaps lower than RNA when tRNA modifications, etc., are considered) and unique chemical identity (although similar to RNA, as represented by the close colors).

A slightly more sophisticated model of the early evolution of nucleic acids offered by Cleaves and Bada [32] hypothesizes some initial emergence of a proto-nucleic acid, followed by template-directed synthesis of RNA from this proto-nucleic acid, or from some other pre-RNA descended from the proto-nucleic acid (Figure 7.2B). Although Cleaves and Bada did not give details on how the transition from one polymer type to another occurred, they note that, in the transition of one polymer type to another, these polymers must be able to form hybrid duplexes. In one simplified interpretation of this model, the polymers are always chemically distinct and chimeric sequences never exist; i.e., the RNA descendent does not result from modification of the ancestor pre-RNA, and is built only by templating from the pre-RNA. DNA is then formed in a similar template-directed manner.

The “Grandfather’s Axe” model of the origin of RNA, proposed by Hud and coworkers [7], is similar to the model by Cleaves and Bada, but gives more detail on how the transition from one polymer type to another occurred (Figure 7.2C). In this model, one distinct type of pre-RNA evolves into another through substitution of one or more of its components. In this scenario, intermediate chimeric sequences in which components of the

ancestor a descendent nucleic acid are present in a single polymer are allowed (although not explicitly required). For this transition to occur, the nucleic acid may briefly transition through a period of higher heterogeneity when exchanging one component for another.

The simplified Cleaves-Bada model and the Grandfather's Axe model conveniently demonstrate two distinct mechanisms of transition from one polymer type to another. In the former case, the descendent polymer is formed intact and separate from the ancestor polymer (but still in a template-directed fashion). In the Grandfather's Axe model, the distinction between one polymer type and another is temporarily blurred, as chimeric sequences of higher heterogeneity are formed, which eventually fully transition to the descendent nucleic acid type. It is not clear to what extent either of these mechanisms operated in the early evolution of nucleic acids.

In many published models of the origin of the nucleic acids, whether by the spontaneous prebiotic emergence of RNA, or through the emergence of some proto-nucleic acid, it is often assumed that the genetic polymer emerged fully intact with low chemical heterogeneity. This was not necessarily the case. However a genetic polymer emerged, it is possible that, it its initial stages, it presented as an irregular, highly heterogeneous polymer, and that some process of refinement, perhaps by a non-Darwinian mechanism of chemical evolution, brought it to a state of lower heterogeneity where template-directed replication could begin. As an example of such a chemical mechanism, polymers that self-assemble in solution tend to be more resistant to hydrolysis, and a polymer with a more regular structure might have a greater chance of self-assembling than an irregular polymer. Because the irregular polymers self-assemble poorly, they are more readily hydrolyzed, and their components are recycled. Eventually, the system accumulates regular polymers

which self-assemble. Because template-directed replication is not emphatically required for this to occur, this process of evolution might be non-Darwinian.

An important note should also be made about adaptive radiation in the early evolution of nucleic acids. If a proto-nucleic acid and pre-RNAs did exist, it is obviously conceptually possible to trace a linear evolutionary sequence of nucleic acids from DNA and RNA back to the proto-nucleic acid. However, starting from the proto-nucleic acid, the lineage that led to RNA and DNA was not necessarily the *only* lineage. Branching in this evolutionary process may have occurred to produce xenonucleic acids (emphatically not pre-RNAs, as they were separate branches that did not evolve into RNA) that have since gone extinct. Although the existence of these ancient xenonucleic acids is impossible to confirm or deny, the principle of adaptive radiation is highly instructive on how RNA might have evolved. As mentioned previously (see § 1.2.1), it is not obvious why ribose, among many possible sugars, was selected. Eschenmoser and coworkers have determined that RNA, when compared to xenonucleic acids formed by different sugars (or different ribose conformations), is optimal in its function [28]. Perhaps the selection of ribose was a truly Darwinian phenomenon in which many nucleic acid variants with different sugars exited, but were out-competed by RNA.

By combining these concepts of the emergence of nucleic acids, a trajectory such as the one in Figure 7.2D can be constructed. This hypothetical trajectory is only one of *many* possibilities of the emergence of RNA, demonstrating that our current understanding of the origin and evolution of life is rudimentary at best.

7.4 Towards a Full Account of the Spontaneous Emergence of Proto-Nucleic Acids

The body of work presented here builds a case for the existence of a primordial proto-nucleic acid, an informational polymer with structural components similar to that of RNA, but which formed more easily under plausibly prebiotic conditions. Principles are also elucidated that were originally intended to describe the proto-nucleic acid, but may apply more strongly to evolutionarily intermediate nucleic acids which were descended from the proto-nucleic acid, but preceded RNA. Put together, an argument is presented that RNA, rather than spontaneously emerging from prebiotic chemical processes, is *evolved*.

The components of a proto-nucleic acid do not necessarily resemble those of RNA in their chemical structures, but are similar in function. Notably, nucleobases other than the canonical set in RNA (adenine, uracil, guanine, and cytosine), and backbones that feature chemical linkages other than the phosphodiester, and trifunctional moieties other than ribose (and possibly compound classes other than sugars), are permitted.

Nucleobases that facilitate self-assembly of monomers or short by complementary association are considered strong candidates. These include, but are not limited to, 2,4,6-triaminopyrimidine (TAP), melamine, 2,6-diaminopurine, adenine, barbituric acid, and cyanuric acid. Of these, TAP, melamine, and barbituric acid react with sugars in water to form glycosides (including nucleosides), and may have been important components of pre-RNA.

A strong candidate for the chemical linkage that forms the backbone of the proto-nucleic acid is the ester. The ester is easily formed under plausibly prebiotic conditions from a “dry-down” reaction in which aqueous solutions of compounds featuring hydroxyl

groups and carboxylic acids are dried at elevated temperature to deposit a concentrated film in which these functional groups react. One class of prebiotically relevant compounds that can oligomerize by esterification are α-hydroxy acids, which are formed by cyanohydrin formation from an aldehyde and cyanide, followed by nitrile hydrolysis. When ammonia is present in this reaction mixture, amino acids are also formed by a similar mechanism in the Strecker reaction. Furthermore, when amino acids and hydroxy acids are combined in a dry-down reaction, depsipeptide oligomers are formed which feature both ester and amide linkages. Therefore, if a prebiotic path to nucleobase-functionalized aldehydes exists, then nucleobase-functionalized depsipeptides could be formed prebiotically.

Although only nucleobases that feature divalent nucleophilic sites successfully react with sugars in water to form nucleosides, a wide variety of nucleobases successfully react with Michael acceptors, such as acrolein, as only a monovalent nucleophilic site is required. The products of the reaction of nucleobases with acrolein are varied, but often produce γ-nucleobase-functionalized aldehydes. These aldehydes can act as substrates for the formation of nucleobase-functionalized amino acids and hydroxy acids.

The depsipeptide nucleic acid is the strongest candidate identified for the proto-nucleic acid, both due to its ease of prebiotic formation, and for its ability to permit supramolecular self-assembly. However, many questions still remain about the chemical evolution of proto-nucleic acids. The depsipeptide nucleic acid system studied here was formed from monomers of regular structure: a γ-nucleobase-functionalized hydroxy acid, in the *S* stereochemical configuration, coupled through an amide to aspartic acid, also in the *S* configuration. In a more prebiotically realistic scenario, both amino acids and

hydroxy acids functionalized with nucleobases would be formed, and would be present as racemates. Additionally, racemic hydroxy acids and amino acids other than aspartic acid would be present that would co-oligomerize with the nucleobase-functionalized monomers. The resulting depsipeptide oligomers would be highly irregular in structure. The processes through which these highly heterogeneous polymers were refined is currently unknown and deserves further study.

7.5 References

[1] Crick, F., Central Dogma of Molecular Biology. Nature, 1970. 226: p. 561-563.

[2] Gesteland, R.F., T.R. Cech, and J.F. Atkins, The RNA World. 3rd ed. 2006, Cold Spring harbor, NY: Cold Spring Harbor Laboratory Press. 768.

[3] Joyce, G.F. and L.E. Orgel, Prospects for understanding the origin of the RNA world, in The RNA World, Second Edition: The Nature of Modern RNA Suggests a Prebiotic RNA World, R.F. Gesteland and J.F. Atkins, Editors. 1999, Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY. p. 49-77.

[4] Becker, S., et al., A high-yielding, strictly regioselective prebiotic purine nucleoside formation pathway. Science, 2016. 352(6287): p. 833-836.

[5] Powner, M.W., B. Gerland, and J.D. Sutherland, Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. Nature, 2009. 459(7244): p. 239-242.

[6] Gibard, C., et al., Phosphorylation, oligomerization and self-assembly in water under potential prebiotic conditions. Nat. Chem., 2018. 10(2): p. 212-217.

[7] Hud, N.V., et al., The origin of RNA and 'My Grandfather's Axe'. Chem. Biol., 2013. 20: p. 466-474.

[8] Cafferty, B.J. and N.V. Hud, Was a pyrimidine-pyrimidine base pair the ancestor of Watson-Crick base pairs? Insights from a systematic approach to the origin of RNA. Israel Journal of Chemistry, 2015. 55(8): p. 891-905.

[9] Rios, A.C. and Y. Tor, On the Origin of the Canonical Nucleobases: An Assessment of Selection Pressures across Chemical and Early Biological Evolution. Israel Journal of Chemistry, 2013. 53(6-7): p. 469-483.

[10] Segre, D. and D. Lancet, Mutually catalytic amphiphiles: simulated chemical evolution and implications to exobiology., in Exobiology: Matter, Energy, and Information in the Origin and Evolution of Life in the Universe, Proceedings of the Fifth Triest Conference on Chemical Evolution, J. Chela-Flores and F. Raulin, Editors. 1997, Trieste: Kluwer Academic Publishers. p. 123-131.

[11] Menor-Salvan, C., et al., Synthesis of pyrimidines and triazines in ice: Implications for the prebiotic chemistry of nucleobases. Chemistry-a European Journal, 2009. 15(17): p. 4411-4418.

[12] Callahan, M.P., et al., Carbonaceous meteorites contain a wide range of extraterrestrial nucleobases. Proceedings of the National Academy of Sciences of the United States of America, 2011. 108(34): p. 13995-13998.

[13] Fuller, W.D., R.A. Sanchez, and L.E. Orgel, Studies in prebiotic synthesis. VI. synthesis of purine nucleosides. J. Mol. Biol., 1972. 67(1): p. 25-33.

[14] Fuller, W.D., R.A. Sanchez, and L.E. Orgel, Studies in prebiotic synthesis: VII. solid-state synthesis of purine nucleosides. J. Mol. Evol., 1972. 1: p. 249-57.

[15] Kolb, V.M., J.P. Dworkin, and S.L. Miller, Alternative bases in the RNA world: The prebiotic synthesis of urazole and its ribosides. J. Mol. Evol., 1994. 38: p. 549-557.

[16] Chen, M.C., et al., Spontaneous prebiotic formation of a β-ribofuranoside that self-assembles with a complementary heterocycle. Journal of the American Chemical Society, 2014. 136(15): p. 5640-5646.

[17] Cafferty, B.J., et al., Spontaneous formation and base pairing of plausible prebiotic nucleotides in water. Nature Communications, 2016. 7: p. 11328.

[18] Fialho, D.M., et al., Glycosylation of a model proto-RNA nucleobase with non-ribose sugars: implications for the prebiotic synthesis of nucleosides. Organic & Biomolecular Chemistry, 2018. 16: p. 1263-1271.

[19] Gonzalez, M.A., et al., Facile preparation of C-glycosylbarbiturates and C-glycosylbarbituric acids. Carbohydrate Research, 1986. 158: p. 53-66.

[20] Martinez, M.B., et al., Synthesis and conformational-analysis of C-glycosylbarbiturates. Carbohydrate Research, 1990. 199(2): p. 235-238.

[21] Ma, M. and D. Bong, Determinants of cyanuric acid and melamine assembly in water. Langmuir, 2011. 27(14): p. 8841-53.

[22] Cafferty, B.J., et al., Efficient self-assembly in water of long noncovalent polymers by nucleobase analogues. Journal of the American Chemical Society, 2013. 135(7): p. 2447-2450.

[23] Cafferty, B.J., et al., Ultra-sensitive pH control of supramolecular polymers and hydrogels: pK(a) matching of biomimetic monomers. Chemical Science, 2014. 5(12): p. 4681-4686.

[24] Guseva, E., R.N. Zuckermann, and K.A. Dill, Foldamer hypothesis for the growth and sequence differentiation of prebiotic polymers. Proc. Natl. Acad. Sci. U S A, 2017. 114(36): p. E7460-E7468.

[25] Cleaves, H.J., The Reactions of Nitrogen Heterocycles with Acrolein: Scope and Prebiotic Significance. Astrobiology, 2002. 2(4): p. 403-415.

[26] Rodriguez, L.E., et al., Nitrogen heterocycles form peptide nucleic acid precursors in complex prebiotic mixtures. Sci. Rep., 2019. 9(1): p. 9281.

[27] Kim, H.J. and S.A. Benner, Prebiotic stereoselective synthesis of purine and noncanonical pyrimidine nucleotide from nucleobases and phosphorylated carbohydrates. Proc. Nat.l Acad. Sci. U S A, 2017. 114(43): p. 11315-11320.

[28] Eschenmoser, A., Chemical etiology of nucleic acid structure. Science, 1999. 284(5423): p. 2118-2124.

[29] Pinheiro, V.B., et al., Synthetic genetic polymers capable of heredity and evolution. Science, 2012. 336(6079): p. 341-344.

[30] Sheng, J., et al., Structural insights into the effects of 2 '-5 ' linkages on the RNA duplex. Proceedings of the National Academy of Sciences of the United States of America, 2014. 111(8): p. 3050-3055.

[31] Decker, P., P. Schweer, and R. Pohlmann, Identification of formose sugars, presumable prebiotic metabolites, using capillary gas chromatography/gas chromatography-mass spectroscopy of n-butoxime trifluoroacetates on OV-225. J. Chromatogr., 1982. 244: p. 281-291.

[32] Cleaves, H.J. and J.L. Bada, The Prebiotic Chemistry of Alternative Nucleic Acids, in Genesis - In The Beginning: Precursors of Life, Chemical Models and Early Biological Evolution, J. Seckbach, Editor. 2012, Springer Science+Business Media Dordrecht. p. 3-33.

[33] Forsythe, J.G., et al., Ester-Mediated Amide Bond Formation Driven by Wet-Dry Cycles: A Possible Path to Polypeptides on the Prebiotic Earth. Angewandte Chemie-International Edition, 2015. 54(34): p. 9871-9875.

[34] Mamajanov, I., et al., Ester formation and hydrolysis during wet-dry cycles: generation of far-from-equilibrium polymers in a model prebiotic reaction. Macromolecules, 2014. 47: p. 1334-1343.

[35] Walker, S.I., M.A. Grover, and N.V. Hud, Universal sequence replication, reversible polymerization and early functional biopolymers: A model for the initiation of prebiotic sequence evolution. PLoS ONE, 2012. 7(4): p. e34166.

[36] Orgel, L.E., Prebiotic chemistry and the origin of the RNA world. Critical Reviews in Biochemistry and Molecular Biology, 2004. 39(2): p. 99-123.